UNDERSTANDING FACTORS CONTROLLING GAG AND GAG-POL POLYPROTEIN PRECURSOR PROCESSING IN HIV-1

By

GREGORY BLOOM

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FUFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1998

Dedicated to my wife and family for all their love and support.

ACKNOWLEDGEMENTS

I am extremely grateful for the love and support of my family who have encouraged me in my pursuit of higher education. My parents, Bernard and Sue, have always believed that knowledge enriches life and have strove to instill that principle in all their children. I greatly appreciate my wife Anna for her invaluable support and generosity all throughout my career.

I am deeply indebted to my mentor, Professor Ben Dunn, for providing me a great opportunity to develop my skills as a scientist. He has given me invaluable advice and guidance throughout my graduate work. I would also like to thank the members of my supervisory committee Dr. Maureen Goodenow, Dr. Harry Nick, Dr Nancy Denslow, and Dr. Charles Allen, for their expertise and suggestions that helped me to complete the work presented.

The Dunn laboratory has been a wonderful environment for learning as well as research. I thank Bill Farmarie who has always been an inspiration. I am grateful for the support and friendship of the members of the Dunn lab. I am indebted to Jen, Jeff, Deepa, Brian, Marina, Chetana, Kevin, Su, Becky, Patty, Constanza, Jay, Joe and Mike for making my time in the lab a very enjoyable one.

TABLE OF CONTENTS

page

ACKNOWLEDGEMENTS iii

LIST OF TABLES vii

LIST OF FIGURES viii

ABBREVIATIONS x

ABSTRACT xii

CHAPTERS

1 BACKGROUND AND INTRODUCTION 1

Classification 1
Origins and Discovery 2
Current Infection Statistics 4
Viral Structure 8
Genomic Structure 10
Viral Lifecycle 14
Attachment and Core Entry 14
Reverse Transcription 15
Integration 17
Transcription and Expression of Viral Proteins 18
Viral Assembly and Maturation 20
Protease 22
Structure 22
Function 24
Inhibition of Protease Function 27
Gag-Pol 38

2 MATERIALS AND METHODS 46

Gag-Pol 46

Construction of Gag-Pol Precursor Expression System 46
Patient Samples 48
Cloning of 1.7 kb Region of Gag-Pol into Expression System 49
Construction of Chimeric Gag-Pol Precursors 52
Frameshifting Procedure 52
Sequencing of the Gag-Pol Precursor 54
Expression of the Gag-Pol Precursor 54
Analysis of Precursor Expression 55
Protease 56
Cloning of Protease Region from Gag-Pol 56
Expression 57
Purification 58
Kinetic Analysis 59

3 EXPRESSION AND ANALYSIS OF ORIGINAL VARIANTS 60

Rationale 60
Results 62
Discussion 63

4 KINETIC ANAYSIS OF THE PROTEASE REGION FROM ORIGINAL VARIANTS 72

Rationale 72
Results 73
Discussion 73

5 CHIMERIC CONSTRUCTS 77

Rationale 77
Results 78
Discussion 85

6 GAG-POL MUTAGENESIS 89

Rationale 89
Results 90
Discussion 96

7 SUMMARY AND FUTURE DIRECTIONS 101

Summary 101
Future Directions 108

LIST OF REFERENCES....112

BIOGRAPHICAL SKETCH....124

LIST OF TABLES

Table page

1-1. Listing of the number of persons infected with HIV-1 as of 1996 for eight countries 5

1-2. Current listing of mutations in HIV-1 protease associated with drug resistance 31

3-1. Amino acid sequence of the region defining the C cleavage site for each of the eight variants used in this experiment 69

4-1. Kinetic analysis of gag-pol$_{HXB2}$ and three variants of HIV-1 protease for three substrates 74

5-1. Summary listing of each of the chimeric constructs analyzed in this section, showing the gag and pol regions used in their respective construction 84

LIST OF FIGURES

Figure page

1-1. Leading causes of death among persons 25 to 44 years old in the United States from 1982 to 1992 7

1-2. Structure of the HIV-1 virion with structural and accessory proteins identified 9

1-3. Illustration of the genomic map of the HIV-1 and HIV-2 viruses..... 11

1-4. Illustration of the major steps in the life-cycle of the HIV-1 virus 16

1-5. Crystal structure of HIV-1 protease showing active site residues and major structural regions 23

1-6. Structure of the gag-pol precursor showing the eight cleavage sites and their sequences 40

2-1. Procedure used to move the 1500 bp fragment isolated from infected patients into the pGAG-POL expression plasmid 51

2-2. Diagram of the generation of chimeric constructs illustrating the region of gag-pol exchange between variants 53

3-1. Western blot of three variants: D1.10, D2.21, and MD.03 64

3-2. Diagram of the original gag-pol precursor being analyzed in Figure 3-1 65

3-3. Amino acid sequence of the protease region for each of the eight variants used in this experiment 67

5-1. Western blot analysis of chimeras C1.0, C2.0, and C3.0 79

5-2. Western blot analysis of p110 gag-pol processing for chimeras C4.0 and C5.0 81

5-3. Western blot analysis of p110 gag-pol precursor processing for chimeras C6.0 and C7.0 83

6-1. The amino acid sequence of the gag region encompassed by the Spe I and Bgl II restriction sites is shown 92

6-2. Analysis and comparison of variant MD.03, and an altered MD.03 variant containing a mutated C site 93

6-3. Analysis and comparison of variant MD.03, and an altered MD.03 variant containing a mutated C site and K to R mutation 94

6-4. Analysis and comparison of the original variant MD.03 to that of a MD.03 variant in which mutations have been performed to change its gag region to that of the gag regions seen for variants D2.21 and D2.22 95

ABBREVIATIONS

(x) g	times gravitational constant
amp	ampicillin
ben	N-α-benzoyl
bp	base pairs
d	deoxy
DMSO	dimethysulfoxide
E. coli	*Escherichia coli*
EDTA	ethylenediaminotetraacetic acid
FPLC	fast protein liquid chromatography
gag	group specific antigen
IPTG	isopropylthiogalactopyranoside
kana	kanamycin
k_{cat}	turnover number
k_{cat}/K_m	specificity constant
K_i	inhibition constant
K_m	Michaelis-Menten constant
LB	Luria Broth
MES	2-(4-morpholino)-ethane sulfonic acid
mg	milligram

$MgCl_2$	magnesium chloride
min	minute
ml (s)	milliliter (s)
mM	millimolar
N-	amino-
NaCl	sodium chloride
ng	nanogram
nM	nanomolar
Nph	*p*-nitrophenyalanine
NTP	nucleotide triphosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PBMC	peripheral blood mononuclear cell
pmole	picomole
pol	polymerase gene
PVDF	polyvinylidene difluoride
RPM	revolutions per minute
SDS	sodium dodecyl sulfate
TFA	trifluoroacetic acid
Tricine	N-(Tris- (hydroxmethyl) Methyl) glycine
Tris	tris (hydroxymethyl) aminomethane
V_{max}	maximum velocity
µg	microgram

Abstract of the Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

UNDERSTANDING FACTORS CONTROLLING GAG AND GAG-POL POLYPROTEIN PRECURSOR PROCESSING IN HIV-1

By

Gregory Bloom

December, 1998

Chairman: Ben M. Dunn
Major Department: Biochemistry and Molecular Biology

The majority of the proteins produced by the HIV-1 virus are in the form of polyprotein precursors. These precursors must be processed into the individual proteins used by the virus. The structural proteins used by the virus are encoded within a 55 kDa gag precursor. The enzymatic proteins are produced only as part of a larger 160 kDa precursor that includes the gag and pol regions of the virus. Both of these precursors are processed by an aspartic proteinase, termed HIV-1 protease.

The protease is only active as a homodimer and must therefore dimerize and cleave itself free of the remaining polyprotein precursor in order to process the remaining cleavage sites. In order for the virus to produce active virions, the protease must cleave the gag at the correct time and in the correct positions. Even minor variations in the speed and efficiency of cleavage can lead to virions that are either non-infectious or whose ability to infect other cells are greatly diminished. The mechanisms used by the

protease to ensure the proper processing of the precursors is still not well understood. In fact there are elements outside of the protease itself that may regulate precursor processing.

In order to determine what influencing factors are present outside of the protease, and how those factors interact with protease to control polyprotein processing, an expression system was developed that allowed the processing reactions to be followed over time. This system also allowed for the insertion of individual regions of the gag and/or pol regions to determine their effect on processing.

Using this system, it was shown that three factors affect precursor processing, the protease region itself, elements within the gag domain, and possible concomitant effects between the two regions. Mutations of elements within the gag domain were used to show the influence of these changes on the processing phenotype of the precursor. This information should provide useful insight into the understanding of gag and gag-pol precursor processing in HIV-1.

CHAPTER 1
BACKGROUND AND INTRODUCTION

Classification

All viruses classified as belonging to the Retroviridae family share two common features, the first of which is the use of positive strand RNA as their genetic material. The second and more distinguishing characteristic is the use of DNA as an intermediate in the replication of their genome. To accomplish this all viruses within this family contain an RNA dependent DNA polymerase, termed reverse transcriptase. One unique genus within the Retroviridae family is the Lentivirus.

The Lentivirus genus encompasses a large number of viruses that infect and cause disease in a very diverse number of hosts. All diseases caused by members of the Lentivirus genus share similar biological and clinical characteristics. The biological characteristics of infection are as follows: species specificity, nononcogenicity, cytopathic effects in certain cells, infection of macrophages, and latent or persistent infection in specific cells. The clinical characteristics can include long incubation periods, immune suppression, and involvement of the central nervous system (41,58). Some of the more well known viruses within the genus are EIAV (equine infectious anemia virus), FIV (feline immunodeficiency virus), and SIV (simian immunodeficiency virus). While many members of the Lentivirus genus are well known and some are even important economically, none are as well known or as significant to mankind as HIV (human immunodeficiency virus).

There are two unique strains of HIV, HIV-1 and HIV-2. Both are retroviruses and are infectious, but they differ in their ability to cause disease and in their prevalence. HIV-2 is a minor subtype of HIV-1, which was first identified in West Africa shortly after the discovery of HIV(69). HIV-2 can lead to AIDS (acquired immunodeficiency syndrome) but has a much longer pathogenic course. HIV-1 is the more infectious of the two strains, has a much shorter clinical course and is much more widespread. Because of these factors HIV-1 has had an enormous medical, social, and economical impact around the world. It is for these reasons that almost all research to date has focused on the HIV-1 strain. This dissertation will focus solely on the HIV-1 strain as well, and all future references to HIV in the text will refer to the HIV-1 strain only.

Origins and Discovery

The origin of HIV is yet to be determined, but from recent sequence information the virus appears to have been a SIV (simian immunodeficiency virus) that crossed over into humans in the not so distant past, perhaps as recently as the early 1950s (56,73). Within thirty years of crossing over to mankind, HIV had begun to cause disease.

In 1983 Barre-Sinoussi, Chermann, Montagnier, and associates at the Pasteur Institute in Paris isolated a virus containing reverse transcriptase from the lymph nodes of a patient with persistent lymphadenopathy syndrome (3). Many researchers believed the virus isolated by the Barre-Sinoussi group was a member of the HTLV (human T-cell leukemia virus) family. In support of this belief, a group led by Robert Gallo reported the isolation of HTLV from AIDS patients in the same issue of the journal Science (37), and

some researchers at the time believed that a member of the HTLV family was the etiological agent of AIDS.

There were still inconsistencies, however, with the hypothesis that a virus from the HTLV family was the cause of AIDS. One of the main problems was the fact that the newly isolated virus would not immortalize CD4+ lymphocytes as HTLV did, but would grow to a very high titer and kill the cells. The second inconsistency was the fact that AIDS was commonly reported in hemophiliacs. HTLV is very closely associated with cells, and is not found free in the plasma. If only plasma was given to hemophiliacs in need of Factor VII, then it was unlikely an HTLV-like virus was being transmitted, and therefore unlikely to be the causative agent of AIDS. Although there were some cross-reactions of proteins, namely the p24 protein, between HTLV and the virus (36), the DNA sequence and cytological properties were very different.

Following their initial report in the 1983 issue of Science, the Montagnier group published additional studies showing that the newly isolated virus would grow to a high titer in CD4+ lymphocytes and eventually kill them (71). This was unlike the pattern seen in HTLV infection of CD4+ lymphocytes, in which the cells are not killed and are often immortalized. Gallo's group at the NIH (National Institutes of Health) also published additional studies showing isolation of a virus distinct from viruses previously discovered in the HTLV family, but showing cross-reaction with the p24 core protein. This virus was termed HTLV-III (36). At the same time, Levy et al. also had isolated a virus, termed ARV (AIDS-associated retrovirus), from patients with AIDS and from patients from high risk groups for contracting AIDS (59). All the isolated viruses, LAV (lymphadenopathy

associated virus), HTLV-III from Gallo's group, and ARV shared the same characteristics, rapid growth in PBMCs, and the killing of CD4+ lymphocytes. Furthermore, all three viruses showed limited sequence similarity to HTLV. Because of these distinctions, in 1986 the International Committee on Taxonomy of Viruses gave the AIDS virus a separate name, human immunodeficiency virus, HIV (18).

Current Infection Statistics

Since AIDS was first identified in the early 1980s it has quickly grown to become an international problem. It is already epidemic in many regions of the world. Current estimates place the total number of people infected worldwide at around 20 million, most of whom live in sub-Saharan Africa and Asia (87). Table 1 gives a breakdown of the current numbers of persons infected per country as of 1996. The first row in Table 1 shows the large number of infected individuals in India, a country that until recently had almost no known HIV-1 infections. The total number of HIV infections is projected to rise to around 100 million by the year 2000. Of this number, 20 million will be children. At the current rate of infection, one person will contract HIV every 13 seconds and a person will die from AIDS every nine minutes (67). In the United States, the figures are equally shocking. Currently, 1 in every 250 persons is infected with HIV-1. One in every 100 males is infected, while one in every 800 females is infected. Calculations using these figures place the total number of people in the United States infected with HIV-1 at well over one million, and the total number of newly reported AIDS cases at 600,000 in 1996 (8). In addition to the large numbers of people infected in the United States and around the world, there has been a sharp increase in the rate of infection as well.

Table 1-1 Listing of the number of persons infected with HIV-1 as of 1996 for eight countries.

India	2-3 million
South Africa	1.8 million
Uganda	1.5 million
Nigeria	1.2 million
Zimbabwe	1.2 million
Kenya	1.1 million
USA	900,000
Thailand	800,000

Table courtesy of J.A. Levy (58).

In 1991 the number of newly reported AIDS cases in the United States was 45,000. In 1992 that number had grown to 47,000, and in 1996 over 68,000 new AIDS cases were reported (7). An even more poignant indication of how AIDS and HIV infection are affecting the United States is illustrated in Figure 1-1. This figure shows the total number of deaths from all major factors including cancer, injuries, heart disease, homicide, and suicide, among persons 25 to 44 years of age. As this figure indicates, the number of deaths caused by AIDS has increased very rapidly. In 1982 AIDS caused the fewest deaths within this age group, but by 1992 it had become the number one killer of persons 25 to 44 years of age. AIDS related deaths remained the number one killer within this group until 1996. The decrease in AIDS related death starting in 1996 is attributed to the arrival of potent protease inhibitors (8).

Another alarming trend is the shift in the infection pattern of HIV-1. Early in the AIDS epidemic, the primary source of new infections and reservoir of current infections was homosexual males and intravenous drug users. Although these groups still represent a relatively high number of HIV infection rates and AIDS related deaths, other segments of the population are also showing increasing rates of infection, with women and children showing the greatest increases in the rate of infection since 1992. Currently, 8000 children in the United States and over 1.5 million worldwide are infected (14). Many of the increases in the number of infections worldwide are in underdeveloped countries, and most of the infected persons in these countries are too poor to afford any of the current antiviral drugs. Given these two facts, there is an urgent need for more accessible antivirals, and a desperate need for an effective vaccine.

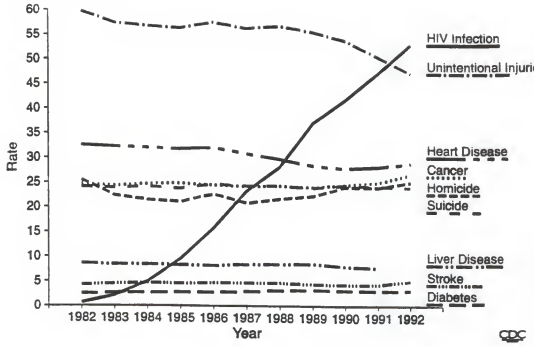


Figure 1-1 Leading causes of death among persons 25 to 44 years old in the United States from 1982 to 1992 (58).

Viral Structure

Figure 1-2 shows the overall structure of the HIV-1 virus. The HIV-1 virus is typical of all Lentiviruses in having a somewhat spherical shape roughly 0.1μm in diameter and a central cone shaped core. The exterior coat of the virus is composed of cellular membrane that was removed upon budding of the virus from the infected cell. This coat is embedded with two major structures derived from the HIV-1 virus. The two proteins are derived from a single precursor, which is later cleaved by a cellular enzyme in the golgi to yield the following two proteins. One protein, gp41 (glycoprotein 41) is a transmembrane protein and plays a role in both fusing of the virion to the cellular membrane and possible incorporation of the preinitiation complex into the cell. The second major structure found on the surface of the virion is the gp120 (glycoprotein 120) protein. This glycoprotein is used by the virus to bind the CD4+ receptor on T-cell lymphocytes (70). These two structures are grouped into 72 knob-like projections on the membrane of the virus. It is thought that each one of the knob-like structures represents a trimer or dimer of each of the proteins (25,38).

The layer directly underneath the lipid bilayer of the membrane is composed of the matrix protein (MA) or p17. This myristoylated core protein has many roles in the life cycle of the virus and is thought to play a major role in maintaining the structural integrity of the virus particle (70). Closely associated with the matrix protein are two proteins Vif and Nef (12,60,109). The function of these proteins will be described shortly. Other viral proteins may also be associated with the p17 core protein such as the viral accessory gene product (Vpr) and Tat, the transactivation protein. Located beneath the p17 layer of

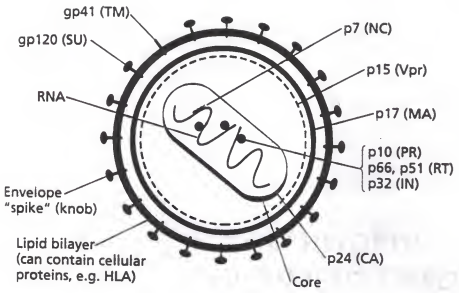


Figure 1-2 Structure of the HIV-1 virion with structural and accessory proteins identified. Adapted from Levy 1998 (58).

protein is the cone shaped core of the virus (62,111). This core is composed of p24 or capsid protein (CA). This core contains two positive strand RNAs that encode the genomic information of the virus. Also contained within the core are the proteins reverse transcriptase RT and protease PR. Two additional proteins, p7 and p6, are also present within the core (70).

Genomic Structure

The HIV-1 genome is approximately 9.8 kb in length. Nineteen proteins are encoded within this genome, many of which are transcribed and translated as larger precursor polyproteins. These polyproteins are posttranslationally cleaved into the individual proteins used by the virus. As can be seen from Figure 1-3, there are several different open reading frames and transframe products within the HIV-1 genome. Located at each end of the genome are the LTRs or long terminal repeats. The LTRs have several functions, the most important of which is encoding of several binding sites for transcription factors and regulators. The gag region is located downstream of the 5' LTR of the genome. The gag region codes for all the structural proteins needed by the virus. The gag proteins are produced either as a p55 gag precursor polyprotein or a p160 gag-pol fusion polyprotein precursor. The proteins encoded by the gag region from 5' to 3' are: p17, the matrix protein; p24, capsid protein, a structural protein; p2, a small spacer protein that plays an important role in the regulation of gag-pol processing; p7 or NC nucleocapsid, a protein which seems to help in reverse transcription; and the p6 protein, associated with budding of the virus particle from the cell (57).

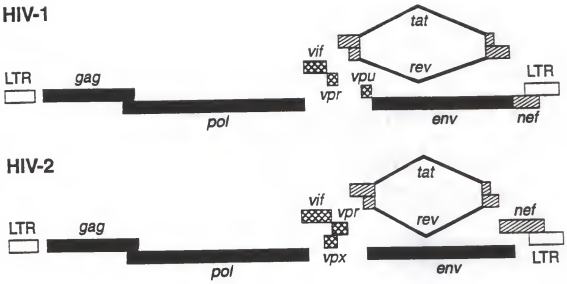


Figure 1-3 Illustration of the genomic map of the HIV-1 and HIV-2 viruses. Note the presence of several overlapping and discontinuous genes. Figure courtesy of J. A. Levy (58).

The region located directly downstream of the gag ORF (open reading frame) is termed pol. The pol region is only produced as part of a larger gag-pol polyprotein precursor termed p160, described previously. This precursor is produced as the result of a –1 ribosomal frameshift near the end of the gag open reading frame. This frameshift occurs only 5 percent of the time. Therefore, 20 times as many structural proteins are produced by the virus than are enzymatic proteins. This mechanism ensures the virus will have enough structural proteins and only the small amounts of the enzymatic proteins encoded by the pol region to survive. The proteins in order from 5' to 3' within the pol region are: protease, PR or p10; reverse transcriptase, RT or p66 and p51; and integrase or p32. PR is responsible for cleavage of the gag and gag-pol precursor polyproteins. Reverse transcriptase is an RNA dependent DNA polymerase that produces dsDNA from ssRNA. Integrase is the protein that catalyzes the insertion of the dsDNA created by RT into the genome of the host cell (57).

The next major downstream region is termed env. Env represents the envelope region. This region codes for the two proteins that function in attachment of the virion to the cell, fusing of the two membranes, and integration of the preinitiation complex. Loops on the gp120 protein bind to regions on the CD4+ receptor on T-cell lymphocytes. Gp41 is the transmembrane protein responsible for fusing and integration. These two proteins are transcribed and translated as gp160 and later cleaved by cellular enzymes contained in the golgi apparatus of the host cell (70). There are several smaller accessory proteins used by the virus encoded either between the pol and env region or as transframe proteins

beginning before the env and ending near the 3' end of the env region (57). The functions of these proteins are described below.

Tat or transactivating protein interacts with a loop structure in the 3' portion of the LTR. Upon binding of tat to this loop, HIV-1 replication is greatly upregulated (76). Rev (regulator of viral expression) protein binds to another loop-like structure in the env region of the mRNA. This binding inhibits the actions of spliceosomes and allows production of full length mRNA. This mRNA is then exported to the cytoplasm and used for packaging into a new virion (31,97). The third major accessory protein is nef. Nef is the least understood of the three main accessory proteins. It appears to down regulate viral expression in some cases and upregulate expression in others. Recently it was discovered that a deletion in the nef gene was the common factor in a large group of long-term nonprogressors. The virus containing the nef deletion was still able to infect the host; however, progression to AIDS has not been observed (91).

There are 3 additional viral accessory genes that serve a variety of functions. Vif causes increased viral infectivity, and cell to cell transmission of the virus. Vpr may also play a role in viral replication, although its function is not yet understood. Vpu is responsible for disruption of gp160 – CD4 interactions and helps in virus release. It should be noted that a large number of these accessory proteins interact with one another to regulate viral expression. Several also require cellular proteins to perform their function (76,81).

Viral Lifecycle

Attachment and Core Entry

Figure 1-4 is a cartoon illustrating the major steps in the life cycle of HIV-1. The first step in the life cycle of HIV is viral attachment to the CD4 receptor of a cell. This attachment is mediated through the actions of the gp120 protein described earlier. The D1 domain, containing the complementarity determining region 2 (CDR2), of the CD4 receptor is the primary area of attachment to gp120, although several areas within the CD4 receptor are likely involved. All cell types expressing the CD4 receptor are capable of becoming infected with HIV-1. These cell types include the primary routes of HIV-1 entry into the body, lymphocytes and macrophages. Although CD4 is the primary receptor for HIV-1, it is alone not sufficient for viral entry into the cell.

In 1997, Berger found that by transfecting CD4 expressing mouse cells, which were resistant to HIV infection, with a protein called fusin, the cells became susceptible to HIV infection (5). Fusin, now termed CXCR4, has subsequently been shown to be a necessary coreceptor for T-cell-line tropic HIV-1 strains. CCR5, a member of the ß-chemokine receptor family, was later discovered to be a coreceptor in the infection of macrophages by macrophage-tropic HIV-1 strains (24). Other cell types, such as human skin fibroblasts, dental pulp fibroblasts, follicular dendritic cells, and brain-derived glial cells, which lack CD4 expression are still infectable by HIV-1. The exact mechanism of infection in these cell lines is not completely understood, but several factors such as complement, the galactosyl ceramide receptor, and MHC may be involved.

Following attachment of the gp120 protein to the CD4 receptor, the core of the virus is injected into the host cell. The factors involved in integration of the core complex are not well defined. It is known that the amino-terminal external portion of the gp41 transmembrane protein is very similar to fusion domains found on paramyxoviruses (35,39). This portion is therefore thought to function in fusing the two membranes after viral attachment. Following fusion of the two membranes, the viral core is injected into the cytoplasm of the cell.

Reverse Transcription

Reverse transcriptase is an RNA dependent DNA polymerase. The function of this enzyme is to produce proviral dsDNA from the +ssRNA carried by the virus. An additional enzyme, Ribonuclease H, along with a cellular tRNA is necessary for this process. The first step in the creation of dsDNA is the reverse transcription of the RNA to produce a negative strand DNA. The cellular tRNA carried by the virus is used as a primer for this synthesis. This newly synthesized DNA and the RNA from the virus form an RNA-DNA hybrid complex. Subsequent to the RNA-DNA hybrid formation, Ribonuclease H degrades the RNA portion of the complex except for an area near the 5' end of the RNA termed the R region, leaving only an ssDNA with a small region of RNA from the virus still complexed at the 5'end. This R region, or possibly the R region of the other +ssRNA carried by the virus, is then used as a primer for the synthesis of the positive strand of DNA. This newly formed dsDNA complex, now called proviral DNA, is then integrated into the host cells genome so that transcription of viral proteins may begin (11,22).

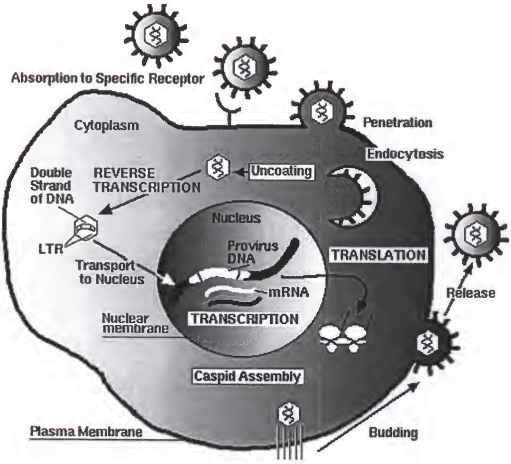


Figure 1-4 Illustration of the major events during the life-cycle of the HIV-1 virus. Figure courtesy of CDC

It is important to note that reverse transcriptase is the sole enzyme responsible for copying the genetic information of the virus. This enzyme is not, however, efficient in its function. The error rate for RT is on the order of 1 misincorporated base pair per 10,000 base pairs in vivo (22). This means that at least one base pair is misincorporated each time the genome is copied. In addition to the low fidelity of the RT, three additional factors contribute to the very large amount of genetic diversity seen in HIV. The first factor is the lack of any proofreading function during copying of the genome. In most cases DNA polymerase enzymes are accompanied by an additional enzyme that corrects any misincorporations during DNA replication. The lack of a proofreading function within HIV does not allow for the errors made by RT to be corrected. The second factor contributing to the high amount of genetic variability is the process of intermolecular strand switching. Although this does not increase the number of mutations within the genome it can greatly affect the amount of genetic diversity. The third reason for the large amount of diversity is the extremely high turnover rate of the virus. It has been shown that the half-life of a virus within an infected cell is less than 1.5 days (9,42,43,80). This fast turnover gives the virus more opportunities to collect mutations during its life cycle. These factors taken together allow the virus many opportunities to collect mutations, and enable HIV to very quickly develop resistance to a variety of inhibitors. These inhibitors and the mutations that cause resistance to them will be discussed later in this dissertation.

Integration

The next step in the life cycle of HIV is integration. Integration is the process whereby the newly formed proviral DNA is imported into the nucleus of the cell and

inserted into the genome of the host cell. This process is very complex and involves a variety of factors. The first step in integration is the circularization of the provirus. Following circularization, the proviral DNA, RT, integrase, Vpr and the p17 matrix protein associate to form the preintegration complex. Two NLS (nuclear localization signals) within the matrix protein are thought to target the preintegration complex to the nuclear pore (101). The nuclear pore is a 90-100 nm structure on the nuclear membrane that controls the movement of molecules in and out of the nucleus. HIV can infect a cell regardless of whether it is actively dividing or at rest, but the preintegration complex can only enter the nuclear pore in actively dividing cells.

After the preinitiation complex enters the nucleus, integrase cleaves the host cell's genome to generate two overhangs each with 5 nucleotides on the 5' strand. Integrase also cleaves the circular provirus and generates two 3' overhangs with a two nucleotide recession. The cleaved provirus then associates with the cleaved genomic DNA, and is ligated by cellular ligase. The 3 base pair gap is subsequently filled in by the DNA repair machinery of the cell (78). Following the completion of integration, expression of the viral proteins begins.

Transcription and Expression of Viral Proteins

The initial mRNA transcripts to be produced following insertion of the proviral DNA are tat, rev, and nef. As described earlier, these proteins are the major regulatory proteins and control the transcription and expression of the HIV-1 genome. The transcription of these genes is regulated in many ways by transcriptional binding sites located within the LTR (long terminal repeat) regions of the inserted genome. The LTR

contains a wide variety of different transcriptional binding sites including both enhancers and promoters (74).

The nef gene is the predominant transcription product produced during the early stages of transcription, representing 80% of all mRNAs. The nef gene product can have pleiotropic effects. It can enhance viral production or maintain the viral genome in the latent state (12,109). The tat gene product is responsible for determining whether HIV-1 will be actively produced or remains in a latent state. Tat controls this activity by binding to the trans-acting response (TAR) element. Binding of tat to the TAR greatly upregulates the production of HIV RNA (6). The third major regulatory protein is rev. Rev also binds to a response element in order to perform its function, the RRE. The rev response element is a loop-like structure in the env region of singly or unspliced mRNA products (65). The binding of rev to the element stabilizes the mRNA and allows it to be transported through the nuclear pore and into the cytoplasm (66,68).

The expression of the viral proteins is very similar to that of cellular proteins. The host cell's ribosomes are used for protein synthesis as well as the charged tRNAs. The most obvious difference between the expression of the host cell proteins and the virally derived proteins is the nature of the mRNA that is being translated. In eukaryotes the mRNA encodes a single gene and thus a single protein is produced during translation of the mRNA. This is not the case in bacterial or viral protein expression. In many cases, as in HIV-1, the mRNA encodes for more than one protein and may include several separate proteins all with separate functions on the same mRNA. This mRNA is termed polycistronic, while an mRNA coding for only one protein is said to be monocistronic.

The majority of all proteins produced by the HIV virus, including all the structural, enzymatic, accessory, and regulatory proteins are produced from polycistronic mRNAs.

The gag proteins are produced in the form of either a 55 kDa polyprotein precursor or as a 160 kDa polyprotein precursor containing the structural proteins of gag and the enzymatic proteins of pol (40). These two different polyprotein precursors are translated from the same mRNA. The production of the 160 kDa precursor is controlled by a –1 ribosomal frameshift. This ribosomal frameshift occurs as a result of secondary structure, in addition to a "slippery" stretch of sequence within the mRNA. The hairpin loop within the mRNA causes the ribosome to stall during translation and the high A (adenosine) content of that region allows the ribosome to slip backward one base pair. This -1 frameshift occurs very near the end of the gag open reading frame (orf). This frameshift causes the stop codon at the end of gag to be out of frame with the new reading frame of the ribosome, and thus allows read through and subsequent translation of the pol region. This frameshift happens in approximately 1 in 20 gag-pol mRNA transcripts allowing the proper amount of pol enzymes to be produced by the virus (86,95).

Viral Assembly and Maturation

The next steps in the HIV life cycle, following production of the necessary viral RNA and proteins, are those of assembly and budding. Viral assembly is a complex and not fully understood event. It is known that the assembly process takes place at the inner cell membrane. The first step in viral assembly is movement of the viral components to the site of assembly. A myristoylation site on the N terminal glycine of the p17 matrix protein is thought to be the main targeting mechanism for both the gag and gag-pol

precursors (95). Along with the gag and gag-pol polyprotein precursors, the two full-length mRNAs are also targeted to the membrane. Association of the mRNA with the nucleocapsid region of the unprocessed polyprotein is the mechanism by which the two mRNAs are moved to the cellular membrane. tRNA is also complexed with the gag-pol precursor and is packaged with the budding virus to serve as a primer template for the next round of reverse transcription.

After the necessary viral components are assembled at the membrane, the process of assembly begins. The p55 gag and the p160 gag-pol precursors oligomerize following attachment to the membrane. The env region undergoes processing by cellular enzymes in the golgi and is transported to the membrane by the cells own transport machinery. The gp41 and gp120 then associate with the membrane. The assembly process places the two ssRNAs in the center of a roughly spherical shaped structure. This structure then begins to bud from the cell.

Budding of the oligomerized gag and gag-pol complexes involves the completion of the spherical structure, begun during assembly. The viral membrane is formed from a small segment of the cellular membrane "pinching off" during budding. The most important aspect of budding is the processing of the polyprotein precursors, gag and gag-pol, into their individual proteins. This process transforms the spherical core of the virus into the condensed core structure seen in mature virions. The viral protease is responsible for the cleavage events during the budding stage of the viral life cycle. Interruption of this process will still allow virions to complete separation from the cell, but the resulting virions are no longer infectious (46).

Protease

Structure

Toh et al. (104) were the first to publish a paper suggesting that protease from HIV-1 is a member of the aspartic protease family of enzymes. This proposal was based on the following evidence. First, a three amino acid repeat, Asp-Thr-Gly, was found to be present in the HIV-1 protease (49). Second, sequence homology to other aspartic proteases such as pepsin was also discovered (102). Third, inhibition of the HIV-1 protease enzyme by pepstatin, a potent inhibitor of the classical aspartic proteases, was demonstrated (96), and finally the observation was made that mutations of the presumed catalytic aspartic acids abolished the activity of the protease (49).

Since the first crystal structure of HIV-1 protease was reported by investigators in 1989, more than 300 crystal structures have been reported of both the bound and unbound form of the inhibitor (110). These structures have shown protease to be a homodimer consisting of two monomers each 99 amino acids in length. The dimer has a two-fold axis of symmetry. Each monomer contributes one part of the catalytic machinery, Asp-Thr-Gly, positions 25, 26, and 27 respectively, to the active site in the dimer. This is unlike the eukaryotic enzymes, wherein both halves of the active site are contained within a single chain of sequence. As in the eukaryotic enzymes, the catalytic triad is stabilized by a series of hydrogen bonds. The active site is located at the bottom of the active site cleft just above the interface of the two dimers. The dimer attachment is by means of noncovalent hydrogen bonding at the N and C terminus of each monomer (110).

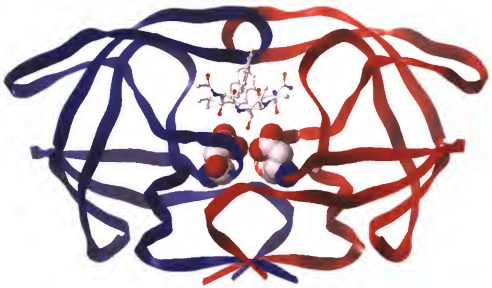


Figure 1-5 Crystal structure of HIV-1 protease showing active site residues and major structural regions.

The active site of the HIV PR is made up of a series of distinct regions lined by amino acids. These regions are known as binding pockets. Each of the binding pockets within the active site is designated with a subsite number, such as S1, S2, S3, etc. The subsites that are located C terminal to the cleavage site are designated S1', S2', S3' etc (92). The group of amino acids that line each binding pocket is unique. Therefore, each binding pocket may have a different three-dimensional structure along with a different energy environment. These two factors contribute to the specificity of each binding pocket of the active site.

Another important feature of the HIV-1 protease is a flap region located at the top of the binding site cleft. This flap region consists of two distinct regions unlike the flap region from the classical enzymes, which is a single structure. This flap region is derived from a β-hairpin loop from each monomer. The flap region must open to allow substrate or inhibitors access to the active site and subsequently close upon substrate binding. The flap region also directly contacts the substrate or inhibitor bound to the active site and is thought to stabilize binding (110).

Function

The HIV-1 protease is produced as part of a 160 kDa polyprotein precursor termed $Pr160^{gag\text{-}pol}$. The protease region of two of the precursors must dimerize so that the protease may form an active site and cleave itself from the larger polyprotein. Following cleavage of each monomer of the protease from the polyprotein, the protease completes the cleavage of the remaining sites within the gag and gag-pol precursors. Mutation of the

active site aspartic acid or the use of potent protease inhibitors results in a molecule incapable of correctly autoprocessing the gag and gag-pol precursors (10,47,54).

There are several factors that affect the activity of the HIV-1 protease and in turn the function of the protease. Among these factors are pH, ionic strength, and the overall amount of protease monomers in the budding virion.

The pH optimum for the enzyme's activity is around 5.5. This value is in accordance with the low pH values needed for enzyme activity in this class of enzymes. There is evidence, however, of processing of gag and gag-pol precursor polyproteins by the protease in the presence of significantly higher pH values (107,114). Many of these studies focused on the ability of PR to free itself from the larger gag-pol precursor, and indeed the conditions in which PR can function as part of the larger precursor may be less stringent than those for the mature enzyme.

The ionic conditions within the budding virion also effect the ability of protease to perform its function. Tyagi et al. tested the ability of protease to cleave the fluorogenic substrate Ben-Arg-Gly-Phe-Pro-MeO-Na over a NaCl and $MgCl_2$ concentration range of 0 to 5 M. The ability of the PR to cleave the compound was greatly reduced when a concentration of 2 M $MgCl_2$ was reached. A similar effect was seen with $CaCl_2$ (107).

Another important factor in protease function is the presence of the correct concentration of PR monomers in the budding virion. It has been demonstrated that the equilibrium dissociation constant (K_d) of the dimeric protease is 0.75 and 3.4 nM at 30 and 37° C, respectively. The rate constant for monomer association is around $4 \times 10^5\ M^{-1}\ s^{-1}$. These values are within the range commonly observed for protein-protein interactions

(23). As the virion is budding from the host cell, the space in which the gag-pol precursors reside becomes smaller. As a result, the relative concentrations of PR containing precursors increase. This increase could lead to a condition in which the dimeric form of the enzyme is favored over the monomeric form in the budding virion. Values such as the pH, ionic strength, and the monomer-dimer equilibrium, affect the formation and activity of the protease. In addition to these factors, however, the overall amount of PR activity and the time at which protease activity occurs are also critical factors in the proper functioning of the protease.

Rose et al. (89) showed that mutation of residues 26 or 28 near the active site of the enzyme decreased the catalytic efficiency (k_{cat}) of the HIV-1 protease 4-fold and 50-fold, respectively. The mutant protease showing the four-fold decrease in activity was still infectious although somewhat less than "wild-type." The mutant with a 50-fold decrease in k_{cat} was no longer able to produce infectious virions despite evidence of gag-pol processing. These results indicate that there is a limiting range of activity in which cleavage of the gag and gag-pol precursors still leads to production of infectious virions. Interestingly, it has also been demonstrated that increased PR activity can lead to production of non-infectious virions. When protease is supplied *in trans* under control of the long terminal repeat with a gag or gag-pol substrate, processing of the gag or gag-pol substrate occurs at a high level, but infectious virions are not produced (2,64). Presumably this increase in the amount of active protease may cause processing of the gag and/or gag-pol precursors to begin prematurely in the cytoplasm of the cell and not in the budding particle, thus leading to virions with malformed core structures. These

studies taken together indicate that protease can have a relative broad range of activity and still cleave the gag and gag-pol precursors. However, there is a much more narrow range in which cleavage of the precursors by protease leads to virions that are infectious.

Inhibition of Protease Function

Because of the essential role the protease enzyme plays in the life cycle of the HIV-1 virus, much effort has focused on the design of inhibitors for this enzyme. There are three major classes of inhibitors for the HIV-1 protease. The first class of inhibitors to be designed and tested were termed peptidomimetic (88). Much of the knowledge gained from designing inhibitors for other aspartic proteases, such as renin and pepsin, was used in the design of this class of inhibitor. Peptidomimetic inhibitors of HIV-1 protease are designed to look-like or mimic a naturally occurring cleavage site within the gag polyprotein precursor. They are inhibitory due to the presence of a non-hydrolyzable, transition-state mimetic isostere between the P1 and P1' residues (105). The observed value of the inhibition constant (K_i) of the peptidomimetic inhibitors is in the subnanomolar range. These inhibitors are also very specific for the HIV-1 protease; however, several problems exist with their use in the treatment of HIV-1 infection. They have a very low bioavailability, are poorly soluble, have poor oral absorption, and are metabolized very quickly. Because of these limitations, a second class of inhibitors was developed.

Using the peptidomimetic inhibitors as ligands, many new crystal structures of the HIV-1 protease were obtained. These structures allowed investigators to see the three-dimensional geometry of the active site of the HIV-1 protease. Observation of the data

from the active site showed it to be symmetrical in nature. Using this information, the next group of inhibitors was developed. This class of inhibitors is called the C2 symmetric based inhibitors (29). These inhibitors are designed to fit the geometry of the active site, and thus not allow entry of the natural substrate for cleavage. The inhibition constants of this class of inhibitors were in the low nanomolar range (29,105).

The third and most recent class of inhibitor to be designed is the cyclic ureas. Members of this class are potent nonpeptidic inhibitors of human immunodeficiency virus (HIV) protease. A fundamental feature of these inhibitors is the cyclic urea carbonyl oxygen that mimics the hydrogen-bonding features of a key structural water molecule. The ability of the carbonyl oxygen in both displacing and mimicking the structural water molecule has been demonstrated through crystallographic studies. These inhibitors are highly selective, have a low molecular weight, and high oral bioavailability (52).

There are currently four inhibitors of HIV-1 protease approved for use in the clinic. They are Saquinavir, Ritonavir, Indinavir, and Nelfinavir. Initial treatment of HIV-1 infections using the above inhibitors employed use of only a single compound. It was found that this type of treatment lead to the development of resistant forms of protease in a very short amount of time, on the order of weeks. It is now standard to prescribe the above inhibitors in combination with two other drugs. The two additional drugs are from a class of inhibitors designed to effect the action of the reverse transcriptase enzyme. There are two types of drugs in this class, each having a different mode of action. They both, however, interfere with the viral life cycle at the same step, that of RNA to DNA conversion.

The first type of drug in this class is termed nucleoside analogs. This class includes the first drug to be approved for the treatment of HIV-1 infection, AZT, along with ddI, ddC, 3TC, and d4T. AZT, clinical name Zidovudine, was first developed for treatment of different types of cancers. It, as well as the other members of this type, interferes with the action of the reverse transcriptase by mimicking one of the nucleosides normally inserted into the growing nucleotide chain. The analog does not, however, allow chain elongation to continue after its insertion due to the absence of a 3' hydroxyl group.

The second type of drug in this class is the non-nucleoside analogs. Members of this group include Nevarapine and Delavirdine. These drugs function by binding to the active site of RT in much the same manner as the protease inhibitors described earlier in this dissertation. They are approved only for use in combination with two additional drugs, however, because of the speed at which resistant forms of reverse transcriptase develop.

It is now common practice to use a protease inhibitor in conjunction with a nucleoside and a nonnucleoside analog. The rational behind this approach is to greatly limit the total amount of virus, and to reduce the rate at which the virus turns over. In doing so it is hoped the opportunities for resistance to develop will be lessened. There are also treatment regimes in which a combination of two or three of the available protease inhibitors is used to treat infection. This type of treatment makes the appearance of resistant protease less likely, because the protease must collect enough of the necessary mutations to ensure resistance to a group of inhibitors and not only a single one.

This method is not always successful due to the similar structure and design of each of the inhibitors. This similarity between different inhibitors allows the protease to collect mutations that confer resistance to all of the prescribed inhibitors. A protease with as few as four mutations has been demonstrated to be resistant to all currently available inhibitors in clinical trials. The four positions found to be mutated were M46I, L63P, V82T, and I84V. Mutations M46I and L63P increased the catalytic efficiency of the enzyme between 100 and 360 percent. These mutations are not near the active site and had no observable effects on enzyme specificity. The two additional mutations V82T, and I84V in contrast affected the ability of the enzyme to bind substrates and inhibitors (94). Taken together these mutations limit the ability of the protease to bind the inhibitors and native substrate, but not to a point where precursor processing stops. The increased catalytic efficiency of the enzyme then compensates for any limitation of binding by increasing the rate at which bound substrate is cleaved. Additional studies have shown this and similar effects, and currently a multitude of mutations which confer resistance have been documented (15,20,21,90,94,103). Table 1.2 lists the four major protease inhibitors and all the currently known positions within the protease where mutations that confer resistance appear.

While it is important to document the number and type of mutations and to which compound they confer resistance, it is equally important to understand how resistance is acquired and where the reservoirs of resistance are located during each stage of treatment. Using the triple combination therapy described above, two phases for HIV clearance from the blood were discovered. The first phase lasts for approximately two weeks

Table 1-2 Current listing of mutations in HIV-1 protease associated with drug resistance.

Compound	Amino Acid Change
Saquinavir	L10I/V/R
	M46I/L
	G48V
	I54V
	L63P
	A71V/T
	V82A/F/T
	I84V
	L90M
Ritonavir	K20R
	L33F
	M36I
	M46I/L
	I54V
	L63P
	A71V/T
	V82F
	V82T
	I84V
	L90M
Indinavir	L10I/V/R
	K20M/R/I/L
	L24I
	V32I
	M46I/L
	I54V
	L63P
	I64V
	A71V/T
	V82A/F/T
	I84/V
	L90M
Nelfinavir	D30N
	M36I
	M46I
	A71V
	V77I

following initiation of therapy. Within this first two weeks, the level of circulating virus in the blood declines by more than 99%. The half-life of circulating virus during this phase is around six hours, while the half-life of an infected cell is somewhere around 1.6 days. The second phase is much longer than the first, and is thought to involve only 1% of total virus in the body. During this phase the mean half-life of infected cells is around 16 days (43,79).

Based on initial decay rates, Perelson et. al predicted complete removal of HIV-1 from an infected individual within 2.3 to 3.1 years, providing that the following assumptions were valid (79). The antiretroviral therapy was completely inhibitory over the entire course of therapy. Other compartments were not present in which the half-life of the virus was slower than that for the second phase. All areas containing HIV-1 virus were accessible to the compounds used in treatment, and latently infected cells were not present. Unfortunately many of these assumptions were not valid. Since it is now possible to remove most of the HIV-1 virus from the blood using current combination therapies, much of the effort of HIV-1 treatment research has turned to defining where the remaining compartments or reservoirs of HIV virus are located within the body, and on developing methods to eradicate the virus from these areas.

There are two categories of potential reservoirs of HIV-1 infection within the body; anatomical, and cellular. The main anatomical reservoir is that of the central nervous system (CNS). The involvement of the central nervous system is a common occurrence in the latter stages of HIV-1 infection in both children and adults. This involvement can lead to a number of fatal conditions. Pathology of patients showing CNS

involvement demonstrates macrophage and macrophage-related microglial cells are the primary sites of HIV-1 infection. Sequence analysis of the HIV-1 virus from these cell types has shown the virus to be of the macrophage trophic non-syncytium-inducing phenotype. It has also been demonstrated that the macrophage trophic non-syncytium-inducing phenotype is found in the CNS even if the primary phenotype found in the blood is of the T-cell trophic syncytium-inducing type (28,34,51). This result indicates that the virus located in the CNS was derived from virus found in the macrophages and macrophage-related microglial cells and not from the virus present in the blood. Furthermore, studies have shown the dynamics of virus in the blood and those of virus in the CNS to be independent of one another (26). These findings point to the CNS as a site of HIV-1 virus infection even after elimination of the majority of the virus from the blood. This compartment could then serve as a source of active virus to re-infect other compartments of the body, mainly the blood, after removal of therapy.

Elimination of virus in the CNS reservoir remains problematic. It has been shown that most of the current class of protease inhibitor does not readily cross the blood-brain barrier due to protein binding. The combination of Ritonavir and Saquinavir, however, has demonstrated some effectiveness in suppressing viral levels in the CNS. Suppression of virus replication in the CNS is an important initial first step, but complete elimination is needed to ensure that virus resistant to treatment does not arise. Indinavir has shown the ability to achieve CNS concentrations comparable only to trough plasma values (99). Use of this drug in the treatment of HIV-1 infection could therefore lead to sub-optimal amounts of drug in the CNS, and in turn, to a serious threat of development of resistant

virus. Any movement of virus or cells containing the virus to the blood would then be of great concern. Clearly, the CNS remains an area of great concern if long-term treatment of HIV-1 infection is to be successful.

The other potentially important anatomical reservoir for persons receiving potent antiretroviral therapy is the testes. Specialized tight junctional complexes between Sertoli cells prevent movement of substances from the testicular interstitium to the seminiferous epithelium (93). These complexes also prevent movement of the drugs used in HIV-1 treatment. In addition, other studies have also show the presence of latently infected cells in the secretions of persons infected with HIV-1 (85). In light of these findings, methods of ensuring effective delivery of compounds to this area during retroviral therapy will be important in the complete elimination of HIV from the body.

The second category of reservoir, that of the cellular reservoir, plays an even more important role in sheltering the HIV-1 virus during retroviral therapy. The three potential sites of sanctuary for the virus in the cellular reservoir are the latently infected CD4+ T cells, the macrophages, and follicular dendritic cells (FDSs).

CD4+ T cells can only efficiently be infected with HIV-1 virus in the activated state. The HIV-1 virus can enter an unactivated CD4+ cell, but the pre-integration complex is not integrated into the cellular genome. The stability of the pre-integration complex is limited to hours or days in the cytoplasm of the cell. If the cell becomes activated during this time, integration of the complex will occur. Unlike that for the pre-integration complex, HIV-1 DNA integrated into the genome of a resting CD4+ cell is a stable situation. The presence of the DNA could theoretically last for the lifetime of the

cell, measured in months or years. The number of latently infected CD4+ cells would theoretically be very low, because the HIV-1 DNA must first infect the cell when it is in the activated state, then the cell must subsequently return to the latent state following integration. The ability of the infected CD4+ T cell to survive this process seems to be limited, because the cell must survive the cytopathic effects of the virus and the cytolytic host effector mechanisms long enough for the cell to somehow turn off the transcription of the integrated HIV-1 genome. In spite of these obstacles, latently infected resting CD4+ T cells have been purified (17). The presence of these cells was greatest in persons in the later stages of HIV-1 infection. Additional studies using a quantitative virus culture technique showed that most of these cells contained defective DNA incapable of replication. A small number of these cells, however, harbored replication competent DNA. The number of latently infected CD4+ T cells containing replication competent DNA in the blood was found to be around 5 per 1 million. Additionally, around 7 per 1 million CD4+ T cells containing replication competent DNA where present in the lymph nodes (16). Based on these numbers, it is estimated the number of resting CD4+ T cells carrying replication competent DNA is between 1.5×10^6. This represents an important albeit small pool of persistent viral infection.

Studies have shown that highly aggressive antiretroviral therapy is not effective at removal of the pool of replication competent latently infected cells. Following 30 months of potent therapy, during which time HIV RNA in plasma was undetectable; 0.2 to 16.2 cells per 10^6 resting cells were estimated to carry replication competent HIV-1 DNA (32). Additional studies demonstrate that this pool does not decrease appreciably over time

during the course of potent antiretroviral therapy. Although the existence of a persistently infected pool of CD4 + cells resistant to decrease with therapy is a major concern in treatment, it should be pointed out that in all cases specialized culturing techniques, notably the removal of CD8+ T cells, were necessary to allow this latently infected pool of cells to proliferate in culture (16,17,32). Regardless of this fact, however, strategies have been devised to rid the body of this reservoir. These include simply outlasting the virus by application of long term therapy. This may be difficult given the low level of compliance observed in many patients, and the presence of counter-indications. Other methods of removal include stimulation of this pool by antigens, superantigens, and the use of mitogens. The ability of these methods to eliminate replication competent CD4+ T cells from the body is still being determined.

Macrophages represent another possibly important persistent reservoir of replication competent HIV-1 DNA during the course of antiretroviral therapy. Macrophages are involved in several areas during the course of infection. They are in some cases the primary path of infection and are responsible for the movement of HIV into the central nervous system, the anatomical reservoir of most concern during treatment. The ability of macrophages to harbor latently infected HIV-1 DNA has been demonstrated. They have also been shown to be the source of *de novo* infection following depletion of the CD4+ T cell population (75). Macrophages have been shown to contribute to the production of HIV-1 in the second, slower phase of decay during treatment described earlier in this dissertation. The lifespan of an infected macrophage is around 14 days. This is the same lifespan that HIV-1 virus has in the second phase of

decay during treatment. Latently infected macrophages have been isolated in the spleen and lymph nodes during the course of infection (27). Little data exist on the effect potent retroviral therapy has on the population of macrophages. Without this data the danger this reservoir poses in treatment of the HIV-1 infected individual is unknown. It has been speculated that the macrophage pool will be protected during treatment because of an ATP-dependent transporter, termed P-glycoprotein transporter. This transporter unidirectionally moves certain compounds including all known HIV-1 protease inhibitors across the blood-brain barrier and the gastrointestinal tract. It therefore limits the levels of drug(s) in these areas and thus allows macrophages to survive during treatment (45,48). Furthermore, the low levels of drug(s) present pose a serious risk for resistance virus to develop. The macrophages harboring the resistant virus could in turn infect or reinfect the blood with highly resistant forms of virus.

The third type of reservoir is that of the follicular dendritic cells, or FDCs. These cell form complex networks on the surface of the lymphatic system. There they conjugate large amounts of virus through antigen-antibody complexes. These complexes are attached to the dendritic cells through tight Fc connections (84). It was hypothesized that the virus attached to the surface of the FDCs infects activated CD4+ T cells during their passage through the germinal centers congested with FDC-bound HIV. This suggestion was shown to be possible using an in vivo system in which infection of CD4+ T cells with FDC bound HIV-1 occurred even in the presence of a neutralizing antibody (27). Initial studies indicated that the amount of virus conjugated to the follicular dendritic cells was completely unaffected by either mono or combination antiretroviral therapy

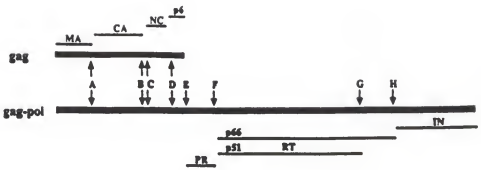
(19). Subsequent studies dispute these findings however, and show a marked decrease in the amount of virus during therapy. Quantitation of the decrease allows for estimates of around one year for the complete elimination of bound HIV-1. The caveat is, however, that even a brief interruption of the treatment regime, or failure to administer at least one protease inhibitor in combination with a reverse transcriptase inhibitor during treatment, can cause a drastic increase in FDC bound HIV-1 virus to pretreatment levels (13). Given the length of the treatment, and the possible side effects, the possibility of a rebound in FDC-bound HIV-1 virus in high. For this reason, much effort has gone into devising a method to stimulate the release of the FDC-bound HIV-1 virus in a more timely manner.

Gag-Pol

In the HIV-1 virus the gag and gag-pol precursors are translated as 55 and 160 kDa precursors, respectively. The protease encoded for within the gag-pol precursor is responsible for cleavage at each of the cleavage sites both within gag and gag-pol. The structure of the gag and gag-pol region is depicted in Figure 1-6. As shown in the figure the processing of the gag and/or gag-pol region occurs at eight distinct sites. The sequence surrounding each of these sites is unique, but the sites themselves can be grouped into three general types or classes. Sites A, E, and F represent class one. Class one sites are distinguished by having an aromatic-Pro at the scissile bond. Class two sites, represented by B, C, and D contain a hydrophobic-hydrophobic sequence at the cleavage junction, and a conserved Gln or Glu residue at the P2' position. The remaining sites, G

and H do not appear to be related to any of the other two classes and have no distinguishing characteristics, and are therefore classed separately.

Several experiments have shown that the order, rate, and timing at which processing of the gag and gag-pol precursors occur is essential for production of infectious virus. Erickson-Vitanen et. al (30) showed the ability of the lysate from an *E. coli* cell transfected with a plasmid producing HIV-1 protease to cleave a full-length gag polyprotein that had been produced *in vitro*. Using immunoblotting, the initial cleavage site was determined to be between the p24-p15 gag proteins (C site). The cleavage at the A site between the p17 and p24 proteins was also observed. The cleavage of these sites could be inhibited with pepstatin A with an IC_{50} of 0.15 mM and 0.02 mM for the p24-p15 and p17-p24 junctions, respectively. Later studies by Krausslich et al used purified HIV-1 protease to cleave both natural substrates produced *in vitro* and synthetic peptides mimicking the gag-pol cleavage sites (50). The enzyme was able to efficiently cleave the *in vitro* produced gag polyprotein precursor at a 20 nM concentration in the assay. Five decapeptide substrates representing the amino acid sequences of the cleavage sites in the gag-pol polyprotein were produced. The purified protease cleaved these substrates with varying rates and efficiencies. The relative order of processing of the synthetic substrates was determined using a competition assay. The order of processing for the sites from first cleaved to last cleaved was E, C, D, A, and B. Subsequent experiments by Pettit et. al (82) using purified HIV-1 protease to cleave a synthetic gag precursor showed that same order in cleavage for the C, D, A, and B sites. The E site was not assayed in this experiment. In addition to the order of processing, the rate at which each of the sites was



b.

TYPE	P_5	P_4	P_3	P_2	P_1		P_1'	P_2'	P_3'	P_4'	LOCATION
	SEQUENCE										
	–Val	Ser	Gln	Asn	Tyr	•	Pro	Ile	Val	Gln–	A
Aromatic*Pro	–Val	Ser	Phe	Asn	Phe	•	Pro	Gln	Ile	Thr–	E
	–Cys	Thr	Leu	Asn	Phe	•	Pro	Ile	Ser	Pro–	F
	–Lys	Ala	Arg	Val	Leu	•	Ala	Glu	Ala	Met–	B
Hydrophobic*	–Ser	Ala	Thr	Ile	Met	•	Met	Gln	Arg	Gly–	C
Hydrophobic	–Arg	Pro	Gly	Asn	Phe	•	Leu	Gln	Ser	Arg–	D
	–Gly	Ala	Glu	Thr	Phe	•	Tyr	Val	Asp	Gly–	G
	–Ile	Arg	Lys	Val	Leu	•	Phe	Leu	Asp	Gly–	H

Figure 1-6 Structure of the gag-pol precursor showing the eight cleavage sites and their sequences (58).

processed relative to the fastest site processed was determined. In this experiment site C (p2-NC) was the first to be processed. The second site was site D (p1-p6). This site was cleaved 9-fold slower than the C site. Cleavage of the A site (p17-p24) was next, and occurred at a rate 9-fold slower than that seen for the C site. The final site to be cleaved in the assay was the (p24-p2) representing site B. This site was cleaved at nearly 400 X slower than the initial site C. Additional studies, as well as those presented later in this dissertation, agree with the order of processing as defined by the experiments described above.

Following investigations showing the varying rates of processing and the sequential order in which the sites within the gag-pol precursor are cleaved, investigators wanted to identify the factors which control these processes. There seem to be several possible mechanisms by which HIV can control the rate and order at which each of the cleavage sites is processed.

The most basic mechanism is the sequence of the cleavage site itself. The type of amino acids for each of the classes of cleavage sites was described previously, and the amino acids at each of these positions have been demonstrated to be critical to the proper cleavage at each of the sites. Mutation of either the P1 or the P1' residues resulted in a drastic change in the cleavage rate and therefore the order in which the site was cleaved (77,83). As stated above, the sequence surrounding each of the cleavage sites is also unique and appears to play an important role in the affinity of the HIV-1 protease for a particular site. A comparison of amino acid sequences from 46 processing sites from 10 different retroviruses showed a sequence preference for positions P4 through P3' (83).

Additionally, Tritch et. al (106) have demonstrated that mutations of residues as many as five amino acids distant from the scissile bond disrupted the rate at which HIV-1 protease processed a given site. Further work by Tritch et. al attempted to convert the p17-24 (A site) into that of the fast-processing Met-Met (C site). The amino acids surrounding the scissile bond were mutated in an expanding fashion on both the P and P' sides of the cleavage junction. Significant cleavage of the synthetic gag precursor containing this mutant (A site) was not observed until four residues on each side of the scissile bond were changed to those found at the (C site) in the LAI sequence. The newly constructed site was still cleaved 10-fold slower than the Met-Met (C site) (106). These and additional data indicate the importance of the residues P4 through P3' in proper cleavage of a given site within the polyprotein precursors, and furthermore demonstrate the influence of additional residues outside of this region that greatly affect processing.

Another factor which has been demonstrated to influence the processing rate of at least one site within the gag and gag-pol polyprotein precursors is the order of cleavage of certain sites within the precursor itself, namely that of the p2-NC site. Blockage of the cleavage of this site was shown to increase by 20-fold the rate at which the p24-p2 site is cleaved. Removal of the region of DNA within the gag domain encoding the p2 protein prior to expression of the synthetic gag precursor also showed a similar rate increase at the p24-p2 site (82). Additionally, virus lacking the p2 protein or virus containing a p2-NC cleavage site not able to be processed still produced virions containing the mature gag cleavage products. Interestingly these virions were much less infectious than virions produced from "wild-type" virus. Unlike that of the p24-p2 cleavage site, cleavage of the

three additional sites within the gag precursor; p17-p24, p1-p6, and p2-NC, occurred at a rate independent of other cleavage events within the precursor.

Another major factor affecting the rate of processing of both the gag and gag-pol precursors is dimerization of the gag-pol precursor itself. The protease encoded within the pol region of the gag-pol polyprotein precursor is active only as a homodimer. Therefore, two gag-pol precursors must dimerize minimally at the area of the pol region containing the protease, so that an active protease can form. Whether this active PR functions as the mature 10 kDa protein or functions as an extended form of the protein containing upstream p6 protein is still uncertain. Studies have shown the ability of the protease to function as an extended protein (1,61,114,115) and this form may indeed be the main type of protease during the processing of the polyprotein precursors. Dimerization of the gag-pol precursors may constitute the rate-limiting step in processing. Studies by Luban et. al (63) demonstrated the effect of gag insertion mutations on dimerization and subsequent processing of the gag and gag-pol precursors. In these experiments 25 linker insertions, four to five amino acids in length, were placed within the matrix, capsid, and nucleocapsid proteins of the gag precursors. The gag-protease construct was expressed in *E. coli* and processing of the products were monitored in a time dependent fashion. The majority of the constructs processed similar to that of "wild-type" gag-pol, however, six of the mutants showed aberrant processing at all of the observed cleavage junctions. When PR was supplied as an additional plasmid, processing of the six mutants was reverted to that observed for the "wild type" construct. Additional studies have shown that mutations and/or deletions within the gag region, mainly within the matrix and capsid

proteins, lead to aberrant processing and release of a noninfectious or less infectious virion (33,44,55,98,108,112).

Taken together these studies demonstrate the ability of changes within the gag region to affect overall polypeptide precursor processing, however the mechanism(s) by which these changes influence processing is still not known. It is possible that mutations in the gag region affect the ability of the gag-pol precursor to dimerize thus preventing protease formation. It is equally as likely however that structural changes caused by mutations in gag disrupt the correct folding of the precursor therefore blocking release of the protease. In addition, mutations in gag may cause additional secondary structure in the gag-pol precursor that may bind to the protease and block access to the active site and thus prevent cleavage of sites within the precursor. Secondary structure in the gag-pol precursor may also cause the disruption in processing observed in the experiments described above by blocking access to the cleavage sites within gag.

As is described above, many possible mechanism(s) exist by which gag may influence precursor processing. In addition, the amount of influence exerted by the gag region on processing may vary greatly depending on the type of mutations found within the gag region. In order to determine to what extent gag influences processing of the gag-pol precursor, and to begin to understand the mechanism(s) by which gag exerts this influence, an *E. coli* based system will be developed in which processing of a gag-pol region isolated from individuals infected with HIV-1 can be followed via Western blot. Using this system several gag-pol regions will be tested and their processing rate and efficiency will be monitored. In addition, chimeric gag-pol precursors will be constructed

in which the gag and pol regions are derived from different precursors. Using these chimeric constructs the role that the individual regions play in overall precursor processing can be further refined.

Additionally, the protease gene from the original set of variants will be cloned, expressed, and purified. Following purification of the protease, kinetic analysis will be used to determine the ability of the protease to cleave a synthetic substrate based on the B cleavage site within the gag precursor. Results from this set of experiments will be used to determine if the values observed in the kinetic assays for the purified protease correlate with the observed processing phenotype seen in the original set of experiments. Following these experiments, site directed mutagenesis will be used to revert mutations found within the gag region of precursors showing an aberrant processing phenotype to those sequences seen in gag regions demonstrating "wild-type" processing. Following the conversion of the residues within the mutant gag region, the processing phenotype of the new precursor will assayed to determine what changes, if any, are present. Using this method it is hoped that those mutations in gag that are responsible for the changes seen in the processing phenotype of the polyprotein precursor can be defined.

CHAPTER 2
MATERIALS AND METHODS

Gag-Pol

Construction of Gag-Pol Precursor Expression System

Two plasmids, pRT/L10 and pDS56-GAG-C6H, were kindly provided by Dr. John Mills of Roche Pharmaceuticals, London (53,54). Plasmid pDS56-GAG-C6H expresses the gag region of HIV-1_{HXB2} under the control of the lac promoter. The pRT/10 expresses the pol region from HIV-1_{HXB2} also under the control of the lac promoter. Both plasmids carry the ampicillin resistance gene, and express their respective proteins in a soluble form. The expression vector M15 was also supplied with the plasmids. The expression system includes an additional plasmid carrying the resistance gene for kanamycin. This plasmid, termed pDMI.1, overproduces the lac repressor allowing for the expression of potentially toxic proteins such as the HIV-1 protease use in these studies.

In order to construct a single plasmid containing the gag-pol region to use as a backbone template for the gag-pol expression experiments described in this dissertation, the two supplied plasmids were digested and ligated together as described below. Twenty-five µl (250ng) of pDS56-GAG-C6H plasmid was added to a microcentrifuge tube containing 2 µl of Xba I (2000 units/µl), 0.4 µl bovine serum albumin (100µg/ml), 8.6 µl H_2O, and 4 µl NEB 2 buffer containing (50nM NaCl, 10 mM Tris-HCl, 10mM

$MgCl_2$, 1 mM dithiothreitol pH 7.9) for a total reaction volume of 40.0 µl. This reaction was incubated at 37°C for 3 hours. Additionally, 25 µl of pRT/L10 (250ng) was incubated with 2 µl KpnI (2000 units/µl), 0.4 µl bovine serum albumin (100 µg/ml), 8.6 µl H_2O, and 4 µl NEB 1 buffer containing 100 mM Bis Tris Propane-HCl, 10mM $MgCl_2$, 1 mM dithiothreitol pH 7.9, at 37°C for 3 hours. Following digestion of the plasmids, 0.5 µl (2.5 units) of the Klenow fragment from DNA Polymerase I and 0.5 µl (2.5 mM) dNTPs were added to both tubes. This reaction was then allowed to proceed for 15 minutes at 30°C. This step was required to fill in the 3' recessed ends of the plasmid created by the previous digestion, in order to obtain the blunt ends needed in the final ligation step of this procedure. At this point, 2 µl of the restriction enzyme Bgl II was added to the reaction containing the now digested and blunt ended pDS56-GAG-C6H plasmid. To the reaction containing the pRT/L10 plasmid, 2 µl of Bam HI in addition to 6 µl of 1 M sodium chloride was added to optimize the buffer for the Bam HI enzyme. Both reactions were incubated at 37°C for an additional 3 hours. The reactions were then electrophoresed on a 1% SeaPlaque™ (FMC) low-melt gel. The digested pDS56-GAG-C6H plasmid, along with the Bam HI – Kpn I fragment from the digested pRT/L10, was excised from the gel using a sterile scalpel. The DNA was purified using the DNA Purification Kit from BIO-RAD in accordance with the manufacture's published protocols.

Following purification, ligation was performed as described below. Six µl of the digested and purified pDS56-GAG-C6H fragment, 9 µl of the Kpn I – Bam HI fragment from pRT/L10, 2 µl of T4 DNA ligase buffer, 1 µl of T4 DNA ligase (6 Weiss units), and

2 µl dd H_2O were combined in a microcentrifuge tube, and the reaction was allowed to proceed for 12 hours at 17°C. The resulting plasmid, now containing a 2400 bp fragment of the gag-pol region between the BamHI and KpnI sites, was subsequently named pGAG-POL. This plasmid will provide the backbone into which the 1.5 kb fragments isolated from the gag-pol region of infected patients will be placed in order for expression to be carried out.

Patient Samples

The region of the gag-pol precursor used in the experiments described in this dissertation was obtained from patients at Shands Teaching Hospital that were infected with the HIV-1 virus. Isolation of the viral DNA from the patients and amplification of the gag-pol regions were performed by colleagues in the lab of Dr. Maureen Goodenow. A description of the isolation procedure performed by members of this lab is as follows. Venous blood, which had been previously been collected in acid citrate dextrose tubes, was centrifuged for 10 minutes at 1400 rpm within six hours of collection. The plasma from these samples was collected and frozen in cyrovials at - 80°C. PBMCs were fractionated by Ficoll-hypaque centrifugation, counted and cryopreserved using a stepfreezer, and stored in liquid nitrogen. DNA was isolated using a salt extraction method (GNOME DNA ISOLATION KIT, Bio101 Inc., La Jolla, CA) and quantified using a spectrophotometer.

Following isolation of the DNA from the PBMCs of patients, a 1.7-kb region extending from the 5' end of gag p24 into reverse transcriptase in pol, was amplified using nested primers and two rounds of amplification. The reactions were performed in a

total volume of 50 µl containing 250 to 500 ng of genomic DNA. In addition to the DNA, each reaction contained 2.25 mM $MgCl_2$, 50 mM KCl, 20 mM Tris (pH 8.2), 100 µg bovine serum albumin per ml, 200 µM each deoxynucleotide triphosphate, 1.25 U taq polymerase, and 100 or 500 nM of each primer used for the first and second round, respectively. The initial round of amplification was performed with the forward primer 5' – GTTAAAAGAGACCATCAAT- 3'. This primer corresponds to nucleotides 935 through 953 in HIV_{LAI}. The reverse primer, corresponding to nucleotides 2710 through 2684 of HIV_{LAI}, was 5' TCCTACATACAAATCATC 3'. The forward primer from the first round of PCR and the reverse primer 5' TTGGAATATTGCTGGTGAT 3' (nucleotides 2614 to 2596) was used for the second round of PCR amplification. A Perkin Elmer 4800 Tempcycler was used for the above reactions. Amplification conditions of the PCR reaction used in this experiment were as follows. Initial denaturation at 95° for 10 min, followed by 35 cycles each consisting of the following conditions: denaturation at 95°C for 1 min, primer annealing at 55°C for 30 seconds, and extension at 72°C for 3 minutes. A final extension at 72°C for 10 minutes was used to ensure complete elongation of products. The sensitivity and specificity of the primers, as well as the accuracy of the Taq polymerase, were assayed using DNA from 8E5 cells. These cells were supplied by the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH.

Cloning of 1.7 kb region of Gag-Pol into Expression System

Following PCR amplification of the 1.7kb region from gag-pol using Taq polymerase, the PCR product was ligated into a pGEM-T vector system™ (Promega

Madison, WI). The pGEM-T vector system takes advantage of the fact that during amplification of DNA using Taq polymerase single 5' adenosine (A) overhangs are generated as a result of the lack of a proofreading function within polymerase. Two thymine (T) nucleotides are located at the 3' end of the pGEM-T vector. The A and T nucleotides anneal and the remaining nick is filled in by T4 DNA ligase. The pGEM-T vector containing the gag-pol region was digested using two unique restriction sites, SpeI and EcoRV, located at the 5' and 3' end of the newly inserted gag-pol region, respectively. The digestion reaction was performed as follows. Thirty µl (300 ng) of pGEM-T containing the gag-pol insert was mixed with 1µl each of SpeI and EcoRV, 0.4 µl 100X BSA, 4 µl of New England Biolabs (NEB) buffer 2, and 13.6 µl of ddH_20. The reaction was incubated at 37°C for approximately 30 minutes. Following digestion, the reaction was mixed with 7 µl of DNA loading dye and separated using electrophoresis on a 1% low-melting agarose gel. Following visualization of the DNA using EtBr, a sterile scalpel was used to excise the band corresponding to the 1.5 kb SpeI – EcoRV fragment. The original Gag-Pol expression construct, described above, was digested in a similar manner.

Following digestion and isolation of the 3.9 kb fragment from the pGAG-POL plasmid, ligation of the 1.5 kb variant gag-pol region into the newly digested pGAG-POL plasmid was performed. (See Figure 2-1) Six µl of the gel-purified 1.5 kb fragment (60 ng) was mixed with 1 µl (10ng) of the gel-purified 3.9 kb pGAG-POL fragment in the presence of 1 µl (6 Weiss units) of T4 DNA ligase, 1 µl T4 of DNA ligase buffer containing 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10mM DTT, 1 mM ATP, 25 µg/ml

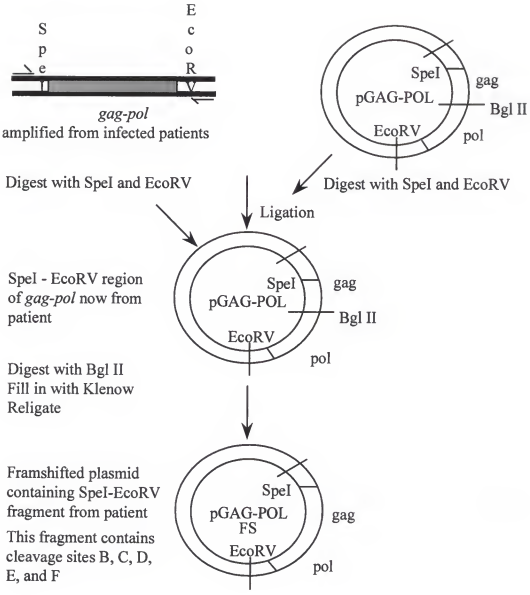


Figure 2-1 Procedure used to move the 1500 bp fragment isolated from infected patients into the pGAG-POL expression plasmid. The frameshifting step is also depicted.

bovine serum albumin, and 1 µl of ddH_20. This reaction was allowed to proceed for 12 hours at 17°C. Following ligation of the variant 1.5 kb gag-pol region into the pGAG-POL plasmid, transformation into the M15 pDMI.1 cell line using the calcium chloride method was performed

Construction of Chimeric Gag-Pol Precursors

Construction of the chimeric Gag-Pol precursors used in this dissertation is shown in Figure 2-2. These constructs were created by first digesting two different naturally-occurring gag-pol regions previously ligated into the pGAG-POL expression vector. Digestion was performed using the restriction enzymes Bgl II and Eco RV. This digestion liberated two distinct bands, one corresponding to the gag region of the gag-pol insert, the other to the pol region along with the remainder of the pGAG-POL plasmid. These products were gel-purified and the gag region from a given variant was ligated with the pol-pGAG-POL region from a different variant. A ligation reaction, similar to the one described previously in this chapter, was used to join these two fragments.

Frameshifting Procedure

In order to express the necessary protease protein from the gag-pol precursor so that processing of the polyprotein could occur, a method to emulate the –1 frameshift that occurs to produce the full-length gag-pol precursor had to be performed. A Bgl II restriction site located near the site of the naturally-occurring frameshift provided a convenient point to perform this manipulation. Digestion of the pGAG-POL plasmid, containing the variant gag-pol region from patient sample, was performed as follows. Twenty µl of plasmid (20ng) was mixed with 1 µl Bgl II restriction enzyme, 2 µl NEB 3

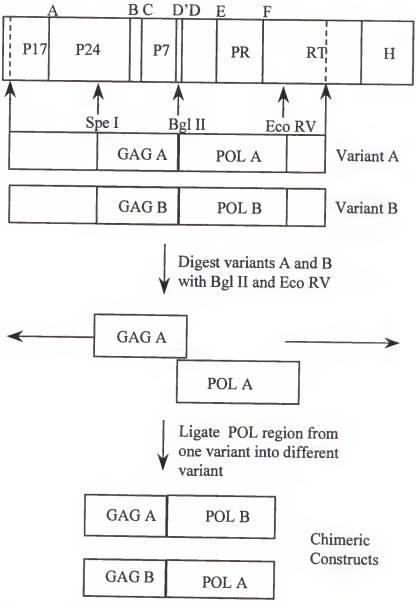


Figure 2-2 Diagram of the generation of chimeric constructs illustrating the region of *gag-pol* exchange between variants. Letters represent cleavage sites within the polyprotein.

buffer (100 mM NaCl, 50 mM Tris-HCl, 10mM $MgCl_2$, 1 mM dithiothreitol pH 7.9), and 7 µl of ddH_2O. The reaction was incubated for 30 minutes at 37°C. Following digestion, the reaction was treated with Klenow to fill in the 3' recessed ends generated by the restriction enzyme. The Klenow reaction was performed as follows. The entire 30 µl of the digest was incubated with 33 µM final concentration of each of the deoxynucleotides along with 1 µl (5 units/µl) of Klenow for 15 minutes at 30°C. The reaction was stopped by heating at 75°C for 10 minutes. This procedure has the effect of adding an additional four base pairs to the Bgl II site causing a –1 frameshift upon read-through by the translational machinery. The inability of the filled in Bgl II site to be cleaved with the Bgl II restriction enzyme was used to confirm the presence of the frameshift.

Sequencing of the gag-pol precursor

Sequence of the 1508 bp region between the restriction sites Spe I and Eco RV was obtained for all the analyzed gag-pol precursors using primers g70, g90, g100, and g110 (72). Generation of the sequences was performed using Sequenase Version 2.0 (U.S. Biochemicals) and an automated ABI 373A sequenator.

Expression of the gag-pol precursor

Following the frameshift procedure, the plasmids were transformed into the M15 *E. coli* cell line. At this point any necessary site directed mutagenesis of residues within the gag region of the polyprotein precursor was performed using the Quick Change ™ kit from Stratagene. The guidelines for design of the necessary primers and the reaction conditions were used as outlined in the accompanying protocol. These cells now contained both the expression plasmid pGAG-POL with the frameshifted, variant gag-pol

region from the patients, and the plasmid containing the lac repressor, pDMI.1. These cells were grown in 50 mls of LB-amp-kana (100 mg/ml amp, 25 μg/ml kana) media until an $O.D._{600}$=0.8 was reached. The cultures were then induced with IPTG (400 μg/ml) Time points were taken at 0, 5, 10, 15, 20, 30, and 60 minutes. At each time point, 300 μl of the culture was removed and an $O.D._{600}$ was taken. An additional 700 μl was also removed and centrifuged at 14,000 rpm for five minutes to pellet the cells. The supernatant was removed and the cell pellet was then lysed with 200 μl of LSB (100 mM Tris-Cl (pH 6.8), 200 mM dithiothreitol, 4% SDS, 0.2% bromophenol blue, 20% glycerol).

Analysis of Precursor Expression

Based on the $O.D._{600}$ readings performed during the expression procedure, equivalent amounts of cells were loaded onto a SDS-PAGE gel (10% separator and 4% stacker Tris-tricine). The samples were electrophoresed for approximately 1 hour at constant voltage. The gels were then blotted onto an Immobilon-P™ (Millipore) PVDF membrane in the presence of 10 mM MES (pH 8.3) buffer for 10-12 hours at 25 milliamps. The electrotransfer was carried out at 10°C. Following the electotransfer, the gels were stained in Coomasie for the presence of remaining proteins. The membranes were used in the Western blot procedure described below.

The membranes were first blocked for 1 hour in the presence of 5% non-fat dried milk in TTBS buffer containing 25 mM Tris (pH 7.4), 100 mM NaCl, 0.05% Tween-20. The membranes were washed briefly with TTBS buffer three times, and once for 15 minutes. The membranes were then incubated with the primary antibody (1 to 10000

dilution) in the TTBS buffer containing 5% non-fat dried milk. The primary antibody was obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH. This monoclonal antibody to p24/25 was developed by K.S. Steimer et. al (100) using the core antigen p25 purified from *E. coli*. This antibody recognizes both the mature p24 capsid protein and all intermediates containing this protein. Following incubation with the primary antibody, the membranes were again washed and subsequently incubated for one hour with the secondary antibody (1 to 3000 dilution). The secondary antibody was a goat anti-mouse antibody conjugated to horseradish peroxidase. The membranes were washed as described above, and an additional thirty minute wash was performed to ensure complete removal of any nonspecific binding. Following the Western blotting procedure, the proteins were visualized using ECL (enhanced chemiluminescence) ™ (Amersham Life Science, Buckinghamshire, England) using the standard protocol described in the accompanying manual.

Protease

Cloning of Protease Region from Gag-Pol

The 99 amino acid HIV-1 protease from several of the variant gag-pol regions from patient samples was cloned into an *E. coli* expression vector so that purified enzyme could be obtained for kinetic characterization. The cloning procedure is described below. PCR was used to amplify a 312 bp fragment containing the gene encoding the PR of HIV-1. The sequences of the primers used were 5' CTCACGGATCCATA-TGCCTGAGATCACTCT 3' for the forward primer, and 5' CTCACGGATCCT-AGCTAGCTAAAAATTTAAAGTGCAACC 3' for the reverse primer. The following

conditions were used for the amplification reaction. An initial denaturation of 94°C for four minutes, denaturation at 94°C for 30 seconds, 30 seconds annealing of the primers at 55°C, and a 30 second elongation step at 72°C. Steps two through four were repeated 25 times. Following amplification the PCR reaction was electrophoresised on a low-melt 1% agarose gel. The band corresponding to the 312 bp fragment was excised with a scalpel and the DNA purified. This fragment was subsequently sequentially digested with NdeI and Bam HI. The purified fragment was isolated and ligated into the pET11a™ (Novagen) expression plasmid. This plasmid expresses protein in the form of insoluble inclusion bodies using a lac/T7 dual promoter system to tightly control expression of potentially toxic proteins. Transformation into the expression cell line BL21DE3 was then performed using the calcium chloride method.

Expression

A 100 ml culture containing M9 minimal media with 5% casamino acids and 100 μg/ml ampicillin was innoculated with a sterile loop of BL21DE3 cells containing peT11a plasmid with the desired insert. This culture was allowed to grow overnight at 37°C. Four 4 liter flasks containing M9 minimal media w/ 5% casamino acids and 100 μg/ml ampicillin were innoculated at 2% and allowed to grow until an $O.D._{600}$ 0.6 – 0.8 was reached. The cultures were induced by the addition of 1mM final concentration of IPTG.

Following induction, 1 ml was removed from the flask at 30, 60, 120, and 180 minutes. An $O.D._{600}$ was taken for each aliquot and the samples pelleted for later analysis on a SDS-PAGE gel. Three hours post-induction the cells were pelleted by centrifugation

at 6,5000 rpm (7,500 X g) for 15 minutes. The supernatant was removed and the cells stored at -20°C for later use.

Purification

The cell pellets were resuspended in TN buffer (Tris Sodium) and lysed using a French pressure cell. The inclusion bodies released from the cells during the lysis procedure were isolated by layering the lysate over a 27% sucrose cushion followed by centrifugation at 8,500 rpm (17,000 X g) for 30 minutes. This process was repeated and the pelleted inclusion bodies were stored at -80°C.

The inclusion bodies obtained from the previous step were solublized in buffer containing 8 M urea, 0.05 M CAPS (pH 10.5), 0.005M EDTA, 0.1 M β-mercaptoethanol. To remove the urea and to refold the protease, dialysis in a buffer containing 0.05 M sodium phosphate (pH 7.3), 0.05 M EDTA, 150 mM NaCl, 0.002 M dithiothreitol was performed. Following dialysis a sample of the refolded protease was tested for activity. Ammonium sulfate precipitation was used to further concentrate and purify the protease. Ammonium sulfate was added to 30, 50 and then 70 % final concentrations. At each of these concentrations, the solution was centrifuged to pellet any proteins that had salted out. A SDS-PAGE gel was used to analyze each of the cuts to determine which contained the protease and to what extent the protein had been purified. Following ammonium sulfate precipitation, the protease was further purified by size exclusion chromatography using a Superdex 75 gel filtration column on a FPLC system.

Kinetic Analysis

The purified protease was used to perform kinetic analysis. Values for K_m, k_{cat}, and k_{cat}/K_m were obtained for four different variants. K_m values were obtained using a diode array spectophotometer (Hewlett Packard). Briefly, a range of concentrations of a synthetic substrate mimicking the Leu*Ala junction at the B cleavage site of the gag region of HIV-1, and containing a ρ-nitrophenylalanine (Nph) at the P1' position as a reporter group, was used to monitor a shift in absorbance from 280 nm to 272 nm. The concentrations of the substrate used in this experiment ranged from 10 to 70 μM. These substrate concentrations were used along with a set amount of enzyme. The tube containing enzyme, and the HIV-1 buffer was incubated three minutes at 37°C prior to addition of substrate, and subsequent monitoring of the reaction. The initial rate of cleavage of the substrate for each concentration was used to determine the kinetic values for each of the purified enzymes.

CHAPTER 3
EXPRESSION AND ANALYSIS OF NATURALLY ARISING VARIANTS

Rationale

The rationale behind construction, expression, and analysis via Western blot, of a region of the gag-pol domain from HIV-1 infected patients was three-fold. The first question to be addressed was whether efficient processing of the gag-pol region would occur when expressed as a single product in an *E. coli* based system. Many researchers have shown the ability of a p55 gag polyprotein precursor, which had previously been expressed in vitro, to be correctly processed by addition of a purified form of the protease from HIV-1. Alternatively, many experiments, including some presented later in this dissertation, tested the ability of a purified HIV-1 protease to cleave a series of synthetic substrates that mimic the naturally-occurring cleavage site found within the gag region. In all of these cases the protease used was supplied in its mature dimeric form and must simply process the substrates provided. The system described here differs from these approaches by forcing the dimerization of PR from two identical gag-pol polyprotein precursors and its subsequent folding into an active conformation. Luban, Goff, and others (63) have shown that dimerization can be the rate-limiting step in processing. This step cannot therefore be overlooked if processing of the gag and gag-pol precursors is to be understood fully. Following dimerization, the PR must cleave itself from the gag-pol precursor in order to complete processing of the remaining cleavage sites. Studies by

Pettit et. al (82) have shown that cleavage of certain sites within the gag region are themselves dependent on previous cleavage reactions. It is conceivable that cleavage of the sites surrounding the protease itself could have similar effects. These effects would only be visible in a system in which protease must first be released from the gag-pol precursor in order to function. For these reasons, this system may more closely mimic the processing events that occur during the budding stage of the life-cycle of the virus, and may therefore provide additional information about the processing of the polyprotein precursors by the protease. The second purpose behind construction of a plasmid to express regions of the gag-pol domain from HIV-1 infected patients was to allow for a convenient method of testing the functionality of a given protease, and indeed a given gag-pol region. Using this method, the ability of a protease to cleave a natural substrate can be accessed in a simple straightforward manner without the need to subclone the protease region. This information can prove useful in making decisions as to the feasibility of a given gag-pol region intended for cloning into a live-virus culture. In addition, this system would provide information at more than one cleavage site per experiment. This is unlike the synthetic substrate systems, which allow values to be obtained for only one site per experiment. Although the information obtained by the cleavage at each of the sites is qualitative in nature, much information about a given protease's ability to cleave its gag region can be ascertained.

The third reason for construction of the pGAG-POL expression plasmid was to provide an efficient method for future testing of individual gag and pol regions within the context of the larger pGAG-POL construct. Manipulation of individual domains within

the gag-pol region should provide valuable information about the roles each of the respective domains plays in the processing of the polyprotein. In addition to determining the contributions to processing of individual domains within gag-pol, this system provides a means to monitor the changes in processing caused as a result of mutating one or more residues within the gag-pol region. This information will be of much value in fully understanding the factors that control the processing of the gag and gag-pol polyprotein precursors.

Results

Products of the expression of the gag-pol region from four variants of HIV-1 infected patients in *E. coli* were analyzed. Figure 3-1 is a Western blot using a p25/24 monoclonal antibody. Variant D1.21 is inactive due to mutation of two residues downstream of the active site aspartic acid. Accumulation of the unprocessed 110 kDa product is seen in lane one. Also visible in lane one are several lower molecular weight bands. These bands are the result of breakdown of the original precursor, possibly by an *E. coli* protease. Lane 2 was loaded with purified p24 capsid protein. This protein serves as a standard marker to determine the size of processed products and to determine if the Western blotting procedure was successful. Lane 3 represents the pre-induction standard. As can be seen, no bands are visible indicating the tight control of the pGAG-POL plasmid carrying the variant gag-pol region prior to its induction with IPTG. The first series of time points 5, 10, 15, 20, and 30 minutes from patient D1.10 were analyzed as described in the materials and methods section of this dissertation. These time points are from patient sample D1.10. As can be seen in all of the lanes for this variant, p25 and or

p25/24 are visible. Variant D2.21is represented in the next set of lanes. For this variant, two bands are visible in the 5, 10 20, and 30 minute time points. The upper band corresponds to the unprocessed gag-pol precursor. The band with a molecular weight of 39 kDa corresponds to a p17-p24 fragment of the gag-pol construct, following processing at cleavage site C. It is important to note that this band would, in cases in which the full-length gag region was used, appear at the 41kDa position; however, in this case, the pGAG-POL plasmid construct lacks the first 2 kDa of the p17 protein. The lower bands are of the p25/24 proteins. The next set of time points is from the MD.03 variant. Here bands corresponding to the p39 protein can be seen for all time points. Bands corresponding to the p25 protein can also be seen for all time points, although much more intensely in the lanes of time points 15, 20, 30, and 60 minutes. Several breakdown products from the 110 kDa precursor, located above the p39, protein can also be seen in the 60 minute lane. In addition to the three variants shown in Figure 3-1, six additional gag-pol variant regions were analyzed using this procedure (data not shown). Each of these constructs exhibited processing essentially identical to one of the constructs shown in Figure 3-1.

Discussion

Figure 3-2 is a diagram representing the region of the pGAG-POL plasmid, into which the gag-pol domain isolated from the infected patients was cloned. Also illustrated are the processing steps used by the protease to liberate the observed fragments seen in Figure 3-1 of the Results section. As can be seen from the results shown in Figure 3-1, and the diagram of the cleavage steps during processing of the variant gag-pol precursor

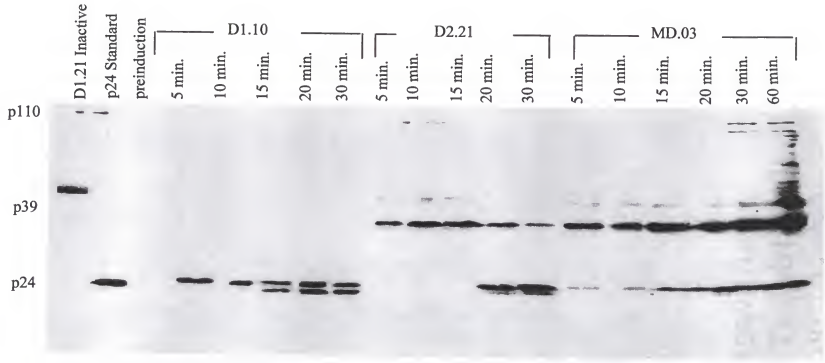


Figure 3-1 Western blot of three variants: D1.10, D2.21, and MD.03. Numbers below each of the variants represents the time in minutes, following induction with IPTG, that the reactions were stopped.

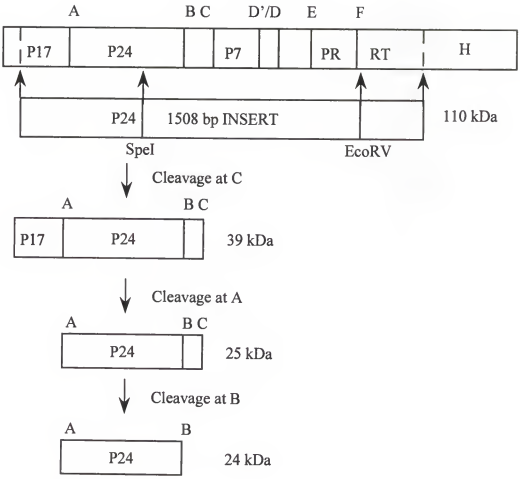


Figure 3-2 Diagram of the original gag-pol precursor being analyzed in Figure 3-1. Each of the Cleavage steps needed by protease to liberate the intermediate products along with the cleaved products themselves is shown.

in Figure 3-2, the variant gag-pol regions produced by the *E. coli* expression system are in all cases processed. The extent of the processing observed in this experiment however, varied greatly among the different variants. These results were unexpected to the extent that the rate and completeness of processing of the variants differed by such large amounts.

The differences in the amount and type of the intermediates observed cannot simply be accounted for by a difference in the amount of sample loaded on the gel because equal amounts of cells were used for each time point among all tested variants. It could also be argued that the amount of a given precursor product produced by the *E. coli* for each of the variants was not equivalent and therefore led to different rates of processing due to the amount of gag-pol precursor available to dimerize and begin the processing procedure. This argument is not, however, supported by the large amounts of unprocessed or partially processed products seen for the D2.21 and MD.03 variants. If the rate-limiting step in processing of the gag-pol precursor is dimerization and this dimerization is solely dependent on the amount of precursor present, it follows that variants D2.21 and MD.03 should have been processed at least as efficiently as the processing seen for variant D1.10. This is clearly not what is observed. From these results it is clear that some factor or factors present in the gag-pol precursor itself, and not factors attributed to experimental analysis, are contributing to the variations in the rate and completeness of processing seen in this set of experiments.

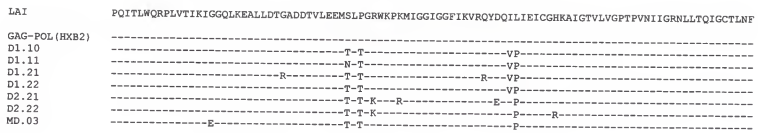


Figure 3-3 Amino acid sequence of the protease region for each of the eight variants used in this experiment.

There are several factors that could be responsible for the large variations observed for the processing of the variants analyzed in Figure 3-1 and the six additional variants analyzed in this experiment. The first of these factors is the sequence of the protease contained within the gag-pol precursor. This is perhaps the most likely factor. Table 3.1 shows the sequence of the protease regions for each of the expressed variants. As can be seen each of the protease sequences for all variants used in this study were indeed different. In addition, research by Leelamanit et. al (manuscript in preparation) has shown that different proteases, which had previously been isolated from patients infected with HIV-1, processed synthetic substrates, based on cleavage sites within the gag region, to different extents. It is therefore possible that the differences in the protease sequence alone are responsible for the variations seen in the processing of these variants. There are, however, additional factors that could influence processing of the gag-pol precursor.

The second such factor is the sequence of the cleavage sites themselves. Barrie et. al (4) sequenced the cleavage sites contained within the gag regions for a large number of HIV-1 infected patients. The majority of sites sequenced in the study were essentially identical to those seen in HIV-1_{LAI}. The cleavage site between p24 and p2 (C site), however, showed large amounts of variation compared to the C site sequence for HIV-1_{LAI}. It is important to note however, that two of the variants used in this set of experiments have identical C cleavage sites (see Table 3-1), but very different processing patterns. This would therefore indicate that the C site sequence, although perhaps important in processing, is not solely responsible for the variations seen in the processing of the gag-pol precursors. The third possible cause of variations observed in processing is sequence variations located outside of the regions defined by the C site and protease. In

Table 3-1 Amino acid sequence of the region defining the C cleavage site for each of the eight variants used in this experiment. The * indicates the site of cleavage by the protease within the sequence.

VARIANT	C SITE
LAI	SATIM*MQRGN
GAG-POL(HXB2)	SQIIM*MQKGN
D1.10	SSIMM*MQRGN
D1.11	SSIMM*MQRGN
D1.21	SQIIM*MQKGS
D1.22	LPNIM*MQKGN
D2.21	SSIMM*MQRGN
D2.22	SSIMM*MQRGN
MD.03	SHLLM*MQRGN

several of the variants used in this experiment, amino acid sequence changes were seen outside of these two regions. These variations could potentially contribute to the processing variations described thus far experiments. These variations and their effect on processing of the gag and gag-pol precursor will be discussed in chapters four and five of this dissertation.

The fourth and final factor, which could account for the differences in the rate and completeness of processing observed for the different variants analyzed in this section, is that of a combination of two or more of the factors previously discussed. Simply put, the overall processing observed for a given variant could be caused by the contributions of two or more of the factors proposed in this section, and not solely be the result of a single element within the gag-pol region. It is also possible that the effect each individual factor contributes to the overall processing phenotype could be negated or enhanced by additional factors within the gag-pol region. For example, the processing phenotype normally produced by a "fast" processing protease could be masked by a C site that has mutated to one that the enzyme cleaves less efficiency. The reverse situation is also possible, in which a "slow" protease produces a phenotype similar to that seen for a faster protease due to elements outside of the protease itself which increase the rate of dimerization or other essential steps essential in cleavage of the gag-pol precursor. This interdependency of each of the factors, and how each of them contribute to the overall processing phenotype, will be the focus of the remainder of this work.

It order to facilitate understanding of the type of processing being referred to in this and future experiments, the entire processing event would best be described in terms

which define both the rate and completeness of the event. To this end, each of the three types of processing observed in Figure 3-1, represented by variants D1.10, D2.21, and MD.03 have been categorized into three distinct classes. Variant D1.10 is described as being fast and complete. The "fast" term is used to define how soon after induction the appearance of the p25/24 protein appears. The second term, "complete", is used to signify whether all of the original 110 kDa gag-pol precursor is processed to the p25/24 protein at one hour post-induction. In variant D1.10 the complete processing of the 110 kDa precursor is seen five minutes after induction, it is therefore classified as "complete". Variant D2.21 is termed intermediate and complete. The term "intermediate" is used due to the later appearance of p25/24 relative to that seen in variant D1.10, and complete by the presence of only the p25/24 protein one hour following induction (not shown). Variant MD.03 represents the third type of class. This class is termed, intermediate, and incomplete. "Intermediate" is used because of the appearance of the p25/24 15 minutes following induction, and incomplete due to the presence of intermediate and unprocessed precursor products one hour post induction. All variants expressed and analyzed in this and all remaining experiments exhibited processing which was describable by one of the classes of processing seen in these three variants. For this reason, this method of classification will be used to describe the results of the expression of gag-pol precursors analyzed by Western blotting for all remaining experiments.

CHAPTER 4
KINETIC ANALYSIS OF THE PROTEASE REGION FROM ORIGINAL VARIANTS

Rationale

It was shown in chapter 3 of this dissertation that processing of the gag-pol regions isolated from patients infected with HIV-1 showed varying rates and completion of processing when analyzed in an *E. coli* expression system by Western blot. Appearance of the lower molecular weight intermediates, p39 and p24, occurred earlier in this set of experiments for some of the variants than others. The p110 protein, representing the entire expressed gag-pol region, was completely processed in some of the variants as early as 5 minutes post induction while others showed accumulation of this precursor form even after 60 minutes post induction. The protease region for all of the variants tested using this system was unique indicating a potential role for the PR in the observed processing phenotypes. The protease within the variant precursors that were processed rapidly could itself be highly active and this activity could therefore account for the rapid rate of observed processing. In those variants showing slow and incomplete processing, a less efficient protease could be the cause. The ability to test the activity of the PR itself within this type of system is limited due to the possible contributions of elements within the gag region of the precursor that could influence processing.

In order to determine if the variations in the processing phenotype seen for the original variants was the result of variations in the activity of the PR within these

precursors, the PR region was cloned, expressed, and purified. The kinetic parameters K_m, k_{cat}, and k_{cat}/K_m were accessed by spectrophotometry, and the resulting values compared to determine if substantial variation in PR activity was present.

Results

The 99 amino acid region representing the PR protein of HIV-1 was expressed, refolded, and purified. Activity of the PR region for four of the variants used in the previous set of experiments, gag-pol$_{HXB2}$, D1.10, D2.21, and MD.03 was analyzed. The results of this analysis are shown in Table 4-1. The top row of the table represents the three substrates mimicking the B cleavage junction within the gag-pol precursor. Three values were obtained for all four variants. The value k_{cat} is the turnover rate of the enzyme with respect to a given substrate. The K_m value is the substrate concentration, given in μM, at which the velocity of the reaction is half maximal. The third value, k_{cat}/K_m, is the specificity constant of the enzyme for a given substrate, and is a good indicator of how well the enzyme functions to cleave a given site. The higher the value of this number, the more efficient the enzyme is at processing the cleavage site. The units for each of the respective values are also shown.

Discussion

As can be seen from Table 4-1 very little variation can be observed in the specificity constant for the four enzymes tested using the HIV-1 substrate. The values for the "fast" protease D1.10 and the "slow" protease MD.03, as defined in the previous chapter, are almost identical. When the HIV-3 substrate was used the values were 0.04 for gag-pol, 0.10 for D1.10, 0.04 for D2.21, and 0.12 for variant MD.03. In this case variants

Table 4-1 Kinetic analysis of gag-pol$_{HXB2}$ and three variants of HIV-1 protease for three substrates designated HIV-1, HIV-3, and HIV-7 is shown. The kinetic values k_{cat}, K_m, and k_{cat}/K_m were obtained for each variant. Standard deviation for the obtained values is also given.

	HIV-1			HIV-3			HIV-7		
	Kcat(s^{-1})	Km (μM)	Kcat/Km ($\mu M^{-1}s^{-1}$)	Kcat(s^{-1})	Km (μM)	Kcat/Km ($\mu M^{-1}s^{-1}$)	Kcat(s^{-1})	Km (μM)	Kcat/Km ($\mu M^{-1}s^{-1}$)
Gag-Pol	0.95 ± 0.08	60.26 ± 7.55	0.02 ± 0.002	1.67 ± 0.18	45.78 ± 6.86	0.04 ± 0.007	1.90 ± 0.88	66.80 ± 8.89	0.03 ± 0.005
D1.10	1.74 ± 0.18	36.22 ± 4.89	0.05 ± 0.008	1.98 ± 0.18	19.33 ± 2.04	0.10 ± 0.01	1.02 ± 0.09	39.28 ± 5.47	0.03 ± 0.004
D2.21	2.47 ± 0.32	72.29 ± 8.89	0.03 ± 0.006	0.76 ± 0.10	19.24 ± 2.85	0.04 ± 0.008	1.21 ± 0.15	17.85 ± 2.54	0.07 ± 0.013
MD.03	1.08 ± 0.11	24.54 ± 3.00	0.04 ± 0.007	1.75 ± 0.17	14.48 ± 1.35	0.12 ± 0.016	1.92 ± 0.23	43.23 ± 6.60	0.04 ± 0.009

D1.10 and MD.03 processed the HIV-3 substrate slightly more efficiently than variants gag-pol and D2.21. Again the "slow" protease showed no less ability to process the substrate then the "fast" protease. In the third panel the substrate HIV-7 was used. Here again very little variation could be observed in processing between the four variants. In this case, D2.21, an "intermediate" protease when tested by the assay described in the previous chapter, was the most efficient at processing the substrate.

There are several conclusions that can be drawn from the results obtained in this set of experiments. First, each of the protease regions tested had the ability to cleave the provided substrate. This indicates that dimerization and activation occurred normally in all of the variants and was therefore not the cause of the variations observed for processing of the gag-pol polyprotein precursor as analyzed by Western blot.

Second, no correlation seems to exist between the ability of any of the proteases tested here to cleave a synthetic substrate and their ability to function in the context of a gag-pol precursor. For example, variant D1.10 showed the fastest processing phenotype using the *E. coli* expression assay. In that assay complete cleavage of the parent precursor, p110, occurred as quickly as five minutes post induction. However, the protease region from that variant does not show an increased "activity" compared to the other enzymes tested in this set of experiments. This indicates the presence of additional factors, perhaps within the gag region of the precursor, that contributed to the observed processing phenotype seen in the first set of experiments. Conversely, variant MD.03 showed incomplete processing of it's gag-pol precursor in the *E. coli* based assay, but when assayed using synthetic substrates showed a slightly better overall ability to process

those peptides compared to the other protease regions tested. This result would again indicate the presence of regions within the gag region of the gag-pol precursor for this variant that, in this case, greatly inhibit the ability of the protease to function as part of it's own gag-pol precursor. These results, in conjunction with the results obtained from the previous set of experiments, indicate that the efficiency of the processing of a given gag-pol precursor is not solely attributable to the efficiency of the protease contained within the precursor, but is greatly affected by additional factors within the gag region, or perhaps the interdependency of factors located in the pol and gag regions of the precursor.

CHAPTER 5
CHIMERIC CONSTUCTS

Rationale

In order to more fully understand the contributions made to processing of the gag-pol precursor by each of the elements contained within the polyprotein itself, a series of gag-pol chimeras were constructed. This construction was described in the Materials and Methods section of this dissertation, Chapter 2. Results from the analysis of the first set of variants left open the possibility that elements in both the gag and pol regions may be responsible for the observed processing phenotypes. In addition, the possibility of contributions by more than one of these factors to the observed processing phenotypes was proposed. These observations made it necessary to develop a system in which the contributions of each of the potential influencing factors could be analyzed.

In addition to understanding the individual roles the gag and pol regions play in the processing of the precursor, it was thought this method would also lead to a greater understanding of the effects any possible interdependency between elements within the gag region and elements within the pol region would have on processing. For example, if changes which had accumulated in the protease and C site regions of the same construct were concomitant, and therefore necessary for the proper processing of the gag-pol precursor, substitution of either the pol region containing the protease or substitution of

the gag region containing the C site, would result in inefficient processing. The changes in processing, if any, would be observable using this system.

In order to ascertain to what extent the factors described above influence processing, a series of chimeric gag-pol precursors were constructed in which a section of the gag region from the gag-pol precursor between the Spe I and Bgl II sites was joined with a section of the gag-pol precursor encompassing the region between the Bgl II and Eco RV restriction sites. This latter region contained the protease enzyme. This construct would therefore allow for the insertion of a gag or pol region from any available variant into the full-length pGAG-POL construct. This new construct could subsequently be expressed and analyzed. Any changes observed in the processing phenotype would then be attributable to the new gag or pol region. This information would therefore demonstrate what effect, if any, different variants of the gag and pol regions have on precursor processing. Using this information, and the sequence of the constructs themselves, mapping of elements within the precursor which influence processing can be addressed. Furthermore, mutation of these elements could allow for further refinement of the role each plays in control of precursor processing.

Results

Products of the expression of chimeric forms of the gag-pol region originating from HIV-1 infected patients were analyzed. Figure 5-1 is a Western blot analysis using a monoclonal antibody for p25/24 protein. The first lane was loaded with variant D1.21. This variant has an inactive protease and therefore accumulates the p110 precursor. Additional breakdown products can also be observed. Lane two is p24 purified protein.

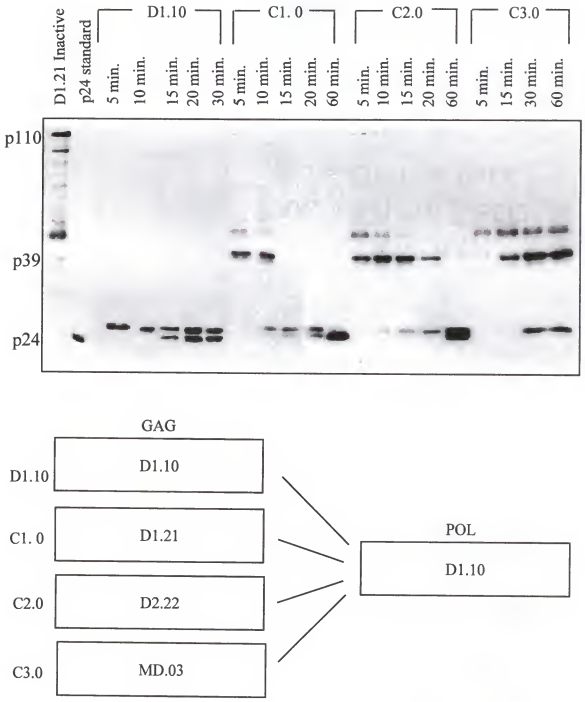


Figure 5-1 Western blot analysis of p110 gag-pol precursor processing for chimeras C1.0, C2.0, and C3.0using a monoclonal antibody against the p25/24 protein. The time points shown represent the duration of processing allowed following induction of expression with IPTG. Also shown are the respective gag and pol regions used for their construction. The original variant D1.10 is shown for reference.

This protein serves as both a marker and an indicator of success of the Western blotting procedure. The next set of five lanes are time points 5, 10, 15, 20, and 30 minutes post-induction for the original variant D1.10. The results of processing for this variant were described in chapter 3. The expression of the C1.0 chimeric variant constitutes the next 5 lanes. In the 5-minute lane a small amount of the p110 precursor, along with p39 and a small amount of breakdown products can be seen. In the next lane, marked 10 min., the same products seen in the 5-minute lane can be seen. There is also an additional band corresponding to the p25 protein. The lanes representing the 15 and 20 minute time-points are identical in constitution, with a very small amount of p39 visible, along with p25 and small amounts of p24. The 60 minute time point shows only completely processed p24 protein. The next set of time points, labeled 5, 10, 15, 20, 50 minutes, are samples for the C2.0 chimera. Time points 5 and 10 minutes show small amounts of p110, breakdown products, and p39. The lanes representing the 15 and 20 minute time points show p39, and small amounts of p25. The 60 minute time point contains only a large amount of p25 and p24. C3.0 is the next variant analyzed. Time points 5 and 15 minutes post-induction contain p39 and varying amount of nonspecific cleavage products. The lanes labeled 30 and 60 minutes contain cleavage products of sizes 39 kDa and 25 kDa, but no p24.

Figure 5-2 represents the next set of chimeras that were analyzed. Variant D2.21 is an original construct and serves as a reference for analyzing the two chimeras shown on the Western blot. The processing of this variant was described previously. Five time

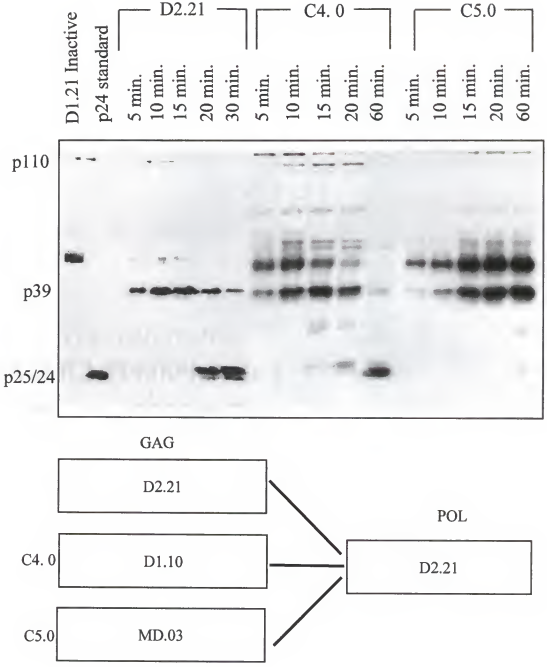
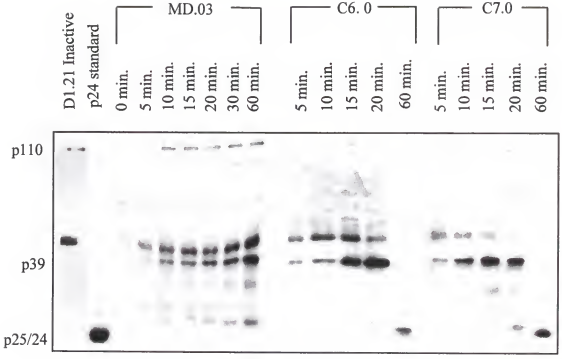


Figure 5-2 Western blot analysis of p110 gag-pol precursor processing for chimeras C4.0 and C5.0. A monoclonal antibody against the p25/24 protein was used in the analysis. The time points shown represent duration of processing allowed following induction of expression with IPTG. Also shown are the respective gag and pol regions used for their construction. The original variant D2.21 is shown for reference.

points corresponding to 5, 10, 15, 20, and 60 minutes post-induction are shown for variant C4.0. Variant C4.0 shows unprocessed p110, breakdown products, and p39 in the lanes representing 5, 10, 15, and 20 minutes post-induction. In addition, small amounts of p25 can be observed in the 15 and 20 minute lanes. The 60 minute post-induction lane shows a very small amount of p39, the remainder of the protein is in the final processed form, p24. C5.0 is the final variant shown on this blot. The same products seen for the 5 through 20 minute time points in chimera C4.0 are also observed for this variant, with the exception of observable amounts of p25/24 in the lanes representing the 15 and 20 minute time points. The lane representing 60 minutes post-induction contains large amounts of p39, and breakdown product, but only trace amounts of p25.

The final set of chimeras analyzed are shown in the Western blot in Figure 5-3. MD.03 is an original variant and was described previously. Chimera C6.0 and chimera C7.0 show very similar patterns of processing, with p25/p24 the only visible protein in the lane representing 60 minutes post-induction.

Table 5-1 summarizes the observed processing phenotype for all of the chimeras analyzed in this set of experiments. The name of each chimera, and the respective gag and pol regions used in its construction are also given. Note, with the exception of the processing phenotype, this information can also be found below the Western blot analysis for each set of chimeras.



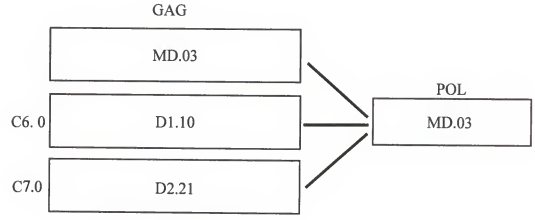


Figure 5-3 Western blot analysis of p110 gag-pol precursor processing for chimeras C6.0and C7.0. A monoclonal antibody against the p25/24 protein was used in the analysis. The points shown represent time in minutes following induction of expression with IPTG that the samples were analyzed. Also shown are the respective gag and pol regions used for their construction. The original variant MD.03 is shown for reference.

Table 5-1 Summary listing of each of the chimeric constructs analyzed in this section, showing the gag and pol regions used in their respective construction. The phenotype of processing observed for each variant is also displayed. *Gag region used in construction for each of these variants was identical.

Chimera	Gag Region	Pol Region	Processing Phenotype
C1.0	D1.21	D1.10	Intermediate/Complete
C2.0	D2.22 (D2.21)*	D1.10	Intermediate/Complete
C3.0	MD.03	D1.10	Slow/Incomplete
C4.0	D1.10	D2.21	Intermediate/Complete
C5.0	MD.03	D2.21	Slow/Incomplete
C6.0	D1.10	MD.03	Intermediate/Complete
C7.0	D2.21	MD.03	Intermediate/Complete

Discussion

The first set of chimeras, C1.0, C2.0, and C3.0 were constructed using the pol region, D1.10. This was done in order to determine if a "fast" processing protease would confer its processing phenotype to a gag-pol precursor irrespective of the gag region used. As can be seen in Figure 5-1, chimeras C1.0 and C2.0, the "fast" processing protease does not convert the gag-pol region(s) it is placed in to it's "fast" phenotype. These results were unexpected, and indicate that elements within the gag region contribute to the observed processing phenotype. Note that the gag regions D1.21 and D2.22 actually decrease the rate at which processing proceeds. Completion of processing does occur however, although the final product of precursor cleavage is not observed until 60 minutes following induction of expression.

Chimera 3.0 also demonstrates an interesting processing phenotype. After 60 minutes of processing there are only very limited amounts of completely processed precursor seen. Indeed, the majority of observed intermediates are in the form of p39 protein. Complete processing of the p39 intermediate was not seen even after allowing the reaction to proceed for three hours following induction of expression (data not shown). This result would indicate that some element or elements within the gag region of variant MD.03 interferes with cleavage at the A site. This cleavage liberates p25/24 from the p39 intermediate. This interference is not easily explained since the A cleavage site in all chimeras are identical in sequence, originating from the HIV_{HXB2} strain used to construct these chimeras. There are at minimum two possibilities that could account for this observation. The first of these possibilities is that cleavage at the C site is somehow

altered in this variant and proper release of the protease following cleavage does not occur. This could prevent cleavage of any remaining sites regardless of the time allowed for the reaction to proceed. An additional possibility is that some as of yet undefined regions within the gag domain of the MD.03 contribute to blockage of the A site following cleavage at site C. Some of the possible factors contributing to this effect will be analyzed in chapter 5 of this dissertation.

Another interesting observation from the set of experiments displayed in Figure 5-1 is that variants D2.22 and MD.03 have very similar processing phenotypes when expressed in their original forms, but very different processing phenotypes when paired with a pol region capable of efficient processing of a precursor polyprotein. D2.22, when paired with the pol region from D1.10, is processed to completion. This result shows that the gag region was not the factor causing the slower/incomplete processing phenotype seen for this variant in the first set of experiments. The opposite conclusion can be drawn for the MD.03 variant however, because pairing of the gag region of MD.03 with any of the pol regions used shows the same, slow/incomplete cleavage seen in the first set of experiments (see Figure 5-1, chimera C3.0 and Figure 5-2, chimera C5.0). In addition, pairing of the pol region of the MD.03 variant with gag regions other than its own allow for efficient, complete processing of the gag-pol precursor, indicating that the pol region of the MD.03 variant is not, in and of itself, defective in processing (see Figure 5-3, chimeras C6.0, and C7.0).

One additional piece of information about the factors that influence gag-pol processing can be obtained from this set of experiments. As evidenced in Table 5-1, all

observed processing for the chimeras, with the exception of those in which the gag domain of variant MD.03 were used, was of the intermediate/complete phenotype. This occurred irrespective of which class of processing phenotype the gag and pol regions from the original set of variants belonged to prior to use in the construction of the chimeras.

For example, it is known from the experiments described in the previous chapter, that the D1.10 pol region has the potential to process its own gag region very quickly and efficiently, as evidenced by the appearance of p25/24 protein only five minutes after induction. However, when the pol region from the D1.10 variant is placed with gag regions other than its own, it was unable to convert these regions to the type of processing observed for this pol variant when used in conjunction with its own gag region. It was, instead, itself converted to intermediate processing. It could be argued that the gag region from the D1.10 variant is simply more efficiently processed, and this fact, and not the efficiency of the pol region, leads to the observed processing. This argument in not, however, consistent with the observations shown in Figure 5-2, chimera C4.0 and Figure 5-3, chimera C6.0. In both of these cases the gag region was processed only at an intermediate rate, with the appearance of the p25/24 protein not discernable until 20 minutes post-induction

These results indicate that in some instances the main factor in determining the rate and completeness at which gag-pol precursor processing occurs is not simply the individual contributions made by the pol or gag domains themselves, but is controlled by interactions between one or more factors from each of the domains. These interactions

could serve to inhibit the rate at which the precursor is cleaved, or as in the case of the D1.10 variant, increase the rate and efficiency of precursor processing. This interdependency could play an important, albeit complex, role in regulation of gag-pol polyprotein precursor processing.

In summary, this set of experiments has provided additional information about the factors involved in the control of processing of the gag-pol precursor. The observation that the gag region does play a role in its own processing, and that protease is not the sole determining factor in this event is novel, and should therefore add to the understanding of gag and gag-pol precursor processing. In addition, the possibility of two or more factors acting in unison seems more likely owing to the large number of chimeras displaying the intermediate/complete processing phenotype observed in this set of experiments, and the observations described for the D1.10 variant.

CHAPTER 6
GAG-POL MUTAGENESIS

Rationale

The results obtained during analysis of the chimeric gag-pol polyprotein precursors indicated that three possible mechanisms exist which can act to influence the processing of a given precursor. They are listed as follows, elements in the pol region of the precursor, elements in the gag region of the precursor, or an interdependent effect of one or more elements in each of the regions. Elements in the pol domain that influence processing are most likely found within the protease region, and several studies describing protease variations that affect cleavage of a synthetic gag precursor have been described previously in this paper. The third possible mechanism responsible for control of gag and gag-pol polyprotein precursor cleavage, that of the interdependency effect seen for variant D1.10, is very interesting and deserves additional study. However, the number of possible variables involved in each experiment, and the number of experiments needed to be undertaken to obtain a suitable answer make study of this factor very difficult. In addition, it is perhaps more beneficial to first define the type of factors present in the gag and pol domains, and to what extent these factors affect processing, before attempting to monitor their combined effects. It is for the reasons stated above that analysis of the factors within the gag region that influence processing will be analyzed in this section.

In addition to these arguments, the results of the chimeric gag-pol expression experiments themselves dictated the next step in the attempt to understand precursor processing. The results indicated that the gag region from the MD.03 precursor was incapable of being cleaved, irrespective of the pol region it was paired with. It was also observed that the MD.03 gag region was unable to be cleaved by its own pol region. In addition, the pol region of MD.03 was able to process gag regions from other variants, indicating its functionality was unimpaired. These results, taken together, strongly indicate that some factor or factors within the gag region of MD.03 itself was the cause of its inability to process. To determine which factor or factors might be the cause of this processing defect, elements in the MD.03 were first mapped by sequencing, and subsequently systematically mutated to elements found in gag regions that show the ability to be processed. It was hoped this method would lead to elucidation of the factors responsible for the observed processing phenotype displayed by the MD.03 variant.

Results

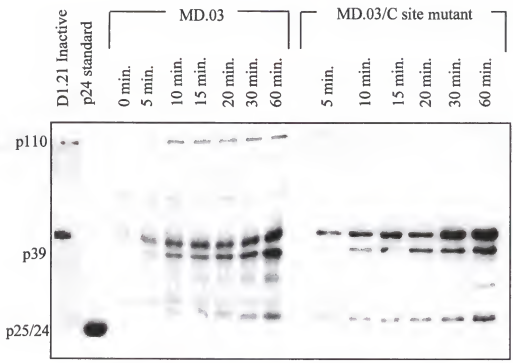
Sequence of the area of the gag domains used in this set of experiments is shown in Table 6-1. This amino acid sequence is located between the Spe I and Bgl II restriction enzyme sites. The amino acid sequence for variants MD.03, D1.10, D2.21, and D2.22 are given. In addition, sequence for two "reference types", one of which, gag-pol $_{HXB2}$, serves as the parental sequence used for this study, is displayed. The amino acids defining the B, C and D' cleavage sites are located within the gray boxes. Residues that differ from the two reference types are in illustrated in white. The locations that correspond to the

positions of the Spe I and Bgl II cleavage sites found in the nucleotide sequence are also displayed.

Figure 6-1 (a) is a Western blot analysis of two variants. MD.03, an original variant from the first set of experiments, is shown in the first panel. The time in minutes, following induction of expression of the gag-pol precursor protein for seven samples are shown in this panel. This processing was described in a previous chapter. Briefly, the 5 through 20-minute time points show only p110 unprocessed gag-pol precursor, products associated with the breakdown of this precursor, and protein corresponding to the p39 intermediate. At the 30 and 60-minute time points, further accumulation of the products described above can be seen, with only relatively small amounts of the p25/24 protein visible. The second panel is a Western blot analysis for the MD.03 variant in which three residues at the C cleavage site have been altered to residues occurring at these sites for the D2.21 and D.22 variants. The sequence of the C site of the MD.03 variant prior to and following mutation of these residues is displayed in Figure 6-1 (b). The processing products obtained 5, 10, 15, 20, 30, and 60 minutes after induction are displayed.

Figure 6-2 (a) is a Western blot analysis of two variants. Again variant MD.03 is from the original set of variants isolated from infected patients, and its processing has been previously described. The second panel represents variant MD.03 in which residues at the C site, and a single K residue downstream of the C site have been mutated to the corresponding residues seen in variants D2.21, and D2.22, see Figure 6-2 (b), prior to its expression and analysis. Samples taken 5, 10, 15, 20, 30, and 60 minutes after inducing expression of the MD.03 variant containing these mutations are shown.

Figure 6-1 The amino acid sequence of the gag region encompassed by the Spe I and Bgl II restriction sites is shown. The sequences for two reference types, HIV_{LAI}, and HIV_{HXB2} are given. Sequence for four variants, including the MD.03 variant sequence, is illustrated. The cleavage sites B, C, and D' used by the HIV-1 protease are defined in gray. Variations of individual residues compared to those of the two reference sequences are shown in white.




```
                Gag                                   Pol
              C-site

 MD.03 ~~~HLL~~~K~~~~~E~~~~~~~~~~~~~~~MD.03
C site ~~~SIM~~~K~~~~~E~~~~~~~~~~~~~~~MD.03
mutant
```

Figure 6-2 Analysis and comparison of variant MD.03, and a variant MD.03 containing an altered C site. a) Western blot analysis of variant MD.03 and variant MD.03 containing mutated C cleavage site using a monoclonal antibody to the p25/24 protein; b) Sequence of the C cleavage site of the original MD.03 variant and the sequence to which the MD.03 site was changed are shown.

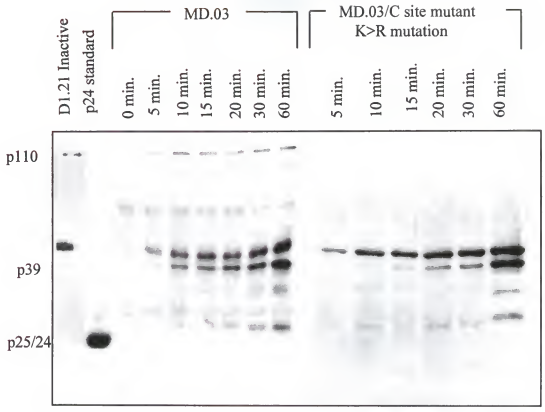
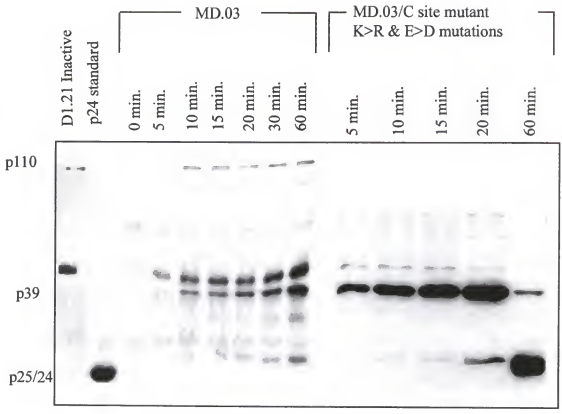


Figure 6-3 Analysis and comparison of the original variant MD.03 and a MD.03 variant containing alterations at its C cleavage site, and an additional mutation N-terminal mutation. a) Western blot analysis of variant MD.03 and the MD.03 variant containing mutations within the C site and an additional K to R mutation using a monoclonal antibody to purified p25/24 protein is shown; b) sequence of MD.03 before and following mutation of the sequence in the regions described above.




```
MD.03 ~~~HLL~~~K~~~~~E~~~~~~~~~~~~~~~MD.03
mutant ~~~SIM~~~R~~~~~D~~~~~~~~~~~~~~~MD.03
C site with
(K>R E>D mutations)
```

Figure 6-4 Analysis and comparison of the original variant MD.03 to that of a MD.03 variant in which mutations have been performed to change its gag region to that of the gag regions observed in variants D2.21, and D2.22. a) Western blot analysis of variant MD.03 and a MD.03 variant containing the mutations described below using a monoclonal antibody to purified p25/25 protein; b) Amino acid sequence of the original MD.03 variant and the sequence modifications used in construction of the new MD.03 variant are illustrated.

Figure 6-3 (a) is a Western blot analysis of two variants. Variant MD.03 has been described above. The second panel represents a mutated variant MD.03 in which residues at the C site, a single K residue, and a E residue have been mutated to the residues seen in the D2.21 variant. Note the appearance of p25 at 20 minutes post induction for this variant compared to the absence of p25/24 in the original MD.03 variant. Also note the almost complete removal of all intermediate products from the lane containing the sample taken 60 minutes post induction indicating the complete processing of the p110 polyprotein precursor.

Discussion

The purpose of this set of experiments was to determine which factors within the gag region of the MD.03 variant contributed to the slow and incomplete processing phenotype observed in the first set of experiments using the naturally arising variants. These results were discussed in chapter 3 of this dissertation. Additional evidence for the presence of elements within the gag region of the MD.03 variant that have the ability to affect processing was obtained from the studies using chimeric gag-pol precursors. The results of these studies are described in chapter 5.

The first step in defining which factor or factors influence processing of the gag region of the MD.03 variant was to obtain sequence information for this region. In addition, sequence information for other variants, whose processing phenotype had been previously described, was also needed in order to compare the sequence found in their gag regions to that of the MD.03 variant. In this manner, it was hoped that sequence changes would be identified in variant MD.03 that were not seen in those variants whose

gag regions were processed more efficiently. Table 6-1 shows the result of this sequence analysis. The sequence information for the two variants D2.21 and MD.03 is of critical importance in determining which factors in the gag region influence processing. As described in chapter 3 and chapter 5 of this dissertation, the variant D2.21 showed an intermediate/complete processing phenotype when coupled with its own and/or other pol regions. These results, therefore, indicate that the sequence of the gag region of the variant D2.21 contains no factors that cause inefficient precursor processing. It then follows that any changes seen in the MD.03 sequence when compared to that of variant D2.21 could potentially be the cause of the aberrant processing seen in the MD.03 variant.

Three areas of difference are seen when comparison of the two sequences is made. The first area of difference is seen at the C cleavage junction. The C site sequence for variant D2.21 is SSIMMMQRGN. The C site sequence for the MD.03 variant is SHLLMMQRGN. The three amino acid changes seen between the two sequences are therefore one possible factor contributing to the differences seen in the processing phenotypes for the two variants. The second and third potential factors are single amino acid changes at residues 7 and 48 amino acids C-terminal of the scissile bond of the C cleavage site. The first change is that of an arginine residue in variant D2.21 to a lysine residue in variant MD.03. The second change is from an aspartic acid in the D2.21 variant to a glutamic acid in the MD.03 variant.

To determine what effect the three amino acid residue changes have on processing, site-directed mutagenesis was used to mutate the three amino acids, H, L, L,

from the MD.03 variant into the three amino acids, S, I, M, seen in the D2.21 variant. No additional changes were made. This new construct was expressed and analyzed as described previously, the result of which is shown in Figure 6-1 (a). As can be seen, the processing phenotypes for the original MD.03 variant and the MD.03 variant containing the mutated C site are identical. In both cases, appearance of p25/24 does not occur until 15 to 20 minutes following induction. More importantly, even after allowing the reaction to proceed for 60 minutes following the induction of expression, the majority of the precursor is found in the intermediate forms. It was therefore concluded that the conversion of the C site in the MD.03 variant to the one found in the D2.21 variant had no discernable effects on processing phenotype.

The second factor that could possibly contribute to the aberrant processing phenotype seen in the MD.03 variant was the R to K change seen seven amino acids from the C cleavage site. Figure 6-2 (a) shows the result of expression of a construct in which both the three amino acids within the C cleavage site and the residue 7 amino acids outside the C site in variant MD.03 were changed to those corresponding residues found in the D2.21 variant. As can be seen from the results for this experiment, no change in the processing phenotype of the gag region for the MD.03 variant occurred, even when these manipulations were performed together.

The third and final factor that could account for the aberrant processing seen in the MD.03 phenotype was the change of the glutamic acid seen in the MD.03 variant to the aspartic acid seen in the D2.21. This MD.03, therefore, has all of the three factors that could potentially be important in determining the processing phenotype converted to the

corresponding factors seen in the D2.21 variant. The results of this experiment are shown in Figure 6-3 (a). As is obvious from the processing phenotype produced, these three changes used in combination did allow the gag region to be processed to completion. Additional studies, in which the C site from the MD.03 variant was left unchanged, and only the K and/or E residues downstream of the C cleavage site were mutated, produced processing phenotypes identical to those observed for the two variants described in Figures 6-1 (a), (panel 1) and 6-2 (a), (panel 1), (data not shown). These results confirm the necessity of all three changes within the gag region of the MD.03 variant in order to convert it into a gag region capable of being processed to completion. There are several possible mechanisms by which these three mutations act in concert to interfere with precursor processing. These mutations may simply prevent dimerization of the gag-pol precursor and thus prevent formation of an active protease. However, the ability of the MD.03 gag-pol precursor to cleave it's own C site indicates that a protease capable of cleavage of at least one site within the polyprotein is formed. This result would indicate that dimerization of the precursor was taking place.

The second mechanism that could lead to the disruption of processing seen in variant MD.03 could be a distortion of the substrate itself so that some of the cleavage sites are no longer accessible. In this scenario the C site could be cleaved initially but any subsequent cleavages may be blocked by other regions of the precursor following this cleavage. However, this scenario is unlikely because studies in which an aliquot of cellular lysate from variant MD.03, 60 minutes post induction, was incubated with increasing amounts of purified PR from the same variant showed that cleavage of the

remaining sites within the precursor was possible, indicating that they were not blocked due to the secondary structure of the substrate itself.

The third possible explanation for the interference of processing observed for variant MD.03 could be attributed to an interaction between the substrate and the protease itself. In this scenario binding of the protease to the mutated C site could lead to a secondary structure within the gag region, perhaps a loop or turn, which could in turn interact either on the surface of the PR, or more likely within or near the active site of the enzyme, and remain attached to the enzyme following cleavage. This interaction could thus prevent the protease from processing any of the remaining sites within the precursor by blocking access to the active site, and would thus result in the processing phenotype observed for variant MD.03. The implications these findings have on understanding the processing of the gag-pol polyprotein precursor will be discussed in the following chapter.

CHAPTER 7
SUMMARY AND FUTURE DIRECTIONS

Summary

The purpose of this work was to further understand the processes that affect the processing of the gag and gag-pol polyprotein precursors by the HIV-1 protease. The work was divided into four areas of research. The first area was determination of the functionality of the processing of a region of the gag-pol precursor when expressed in an *E. coli* system. In order to accomplish this goal, an expression system was designed in which a 1508 bp fragment of the gag-pol region of HIV-1 from infected patients was cloned into the expression vector pGAG-POL. This expression vector produced this variant region in the context of a larger, 2300 bp backbone. This region encompassed all of gag except a 2 kDa fragment at the N-terminal end of the p17 matrix protein. The C-terminal region included the protease gene and a portion of reverse transcriptase. This expression plasmid, under the control of the lac promoter, expressed a 110 kDa protein in a soluble form. After induction of expression of the precursor with IPTG, the level of expression of the precursor and the amount of each of the processing intermediates could be followed by Western blot analysis of SDS-PAGE separated intermediates.

Using the method described above, a series of naturally arising variant gag-pol precursors were analyzed for their ability to be processed in this system. The results of this set of experiments showed that processing of the variant gag-pol precursors did occur

in this system; however, the processing proceeded at varying rates and to different levels of completion. All of the variants used in the first set of experiments demonstrated processing that fit into a defined class or type. The classification of processing phenotypes was used to better describe the phenotype for each variant without the use of long descriptions describing all of the individual events.

The conclusions derived from the first set of experiments were two-fold. First, processing of a variant gag-pol precursor will occur in the designed *E. coli* expression system. Furthermore, the order of processing is similar to that observed using a variety of different methods. The second conclusion drawn from these experiments was that a number of different factors act to control processing of the gag and gag-pol precursors. Three different factors were discussed. The first factor was that of an influence by elements within the pol region of the gag-pol precursor, namely from within the protease itself. Protease is the catalytic enzyme responsible for this processing, and therefore seems the most likely candidate for controlling this event. The presence of this influence is supported by the fact that all of the protease regions for the variants analyzed had different sequences, however, for the nine variants tested only three types of processing phenotypes were observed. This would indicate that protease is not solely responsible for determining the processing phenotype of the precursor.

To determine if there exist a correlation between the observed processing phenotype and the ability of a purified form of the PR to process a synthetic substrate, the protease region from four of the variants tested in the first set of experiments; gag-pol, D1.10, D2.21, and MD.03 was expressed, purified, and assayed. The kinetic parameters

of k_{cat}, K_m, and the specificity constant k_{cat}/K_m were calculated. The results of this work was described in chapter 4. In summary, the ability of a given PR enzyme to cleave a peptide substrate did not seem to correlate in any way with the rate and efficiency at which that PR cleaves the sites within it's own gag-pol precursor. This result supports the conclusion that elements outside of PR have a major influence on the processing of the polypeptide precursor.

The second type of factor, which could affect precursor processing, is the presence of influencing elements within the gag region. This observation was supported by the large number of differences observed within the sequence defining the C cleavage site. The third factor discussed was the possibility that elements in the pol and elements in the gag regions acted together to influence the type of processing seen in the variant polyproteins. These three factors could not be fully tested using the initial system, therefore a method to replace the gag or pol region of one variant with that of another prior to expression was devised. It was hoped this system would provide information about the individual contributions by each of the respective regions, and possibly provide evidence for the interdependency effects proposed following the first set of experiments. The results of the chimeric proteins produced using this system, and the contributions made toward refinement of the factors influencing gag-pol processing, are described below.

As stated earlier, construction of chimeric gag-pol regions was used to monitor the individual influences of elements within the gag and pol regions in overall precursor processing. The results of these experiments, described in chapter 5 of this dissertation,

indicate the presence of elements within the gag region, specifically for variant MD.03, that have a major influence in processing of the precursor. For this variant, the elements present in the gag region caused a major reduction in both the speed and efficiency of processing. This reduction in processing ability was not simply a result of a slow processing protease because pairing of the protease region from this variant with gag regions from other variants resulted in gag-pol polyproteins able to process to completion. Furthermore, the pairing of this variant with any other pol region resulted in polyproteins unable to process completely. These results clearly indicate the presence of factors within the gag region itself that play a major role in interrupting the processing of this precursor.

In addition, it was observed for variant D1.10 that pairing of the pol region from this variant with gag regions from other variants did not impart the "fast" processing phenotype observed for the D1.10 variant. The gag region of this variant was also paired with pol regions from additional variants resulting in processing that was only intermediate. These results, taken together, indicate the presence of factors within the gag and pol regions from this variant that act in conjunction to process the polyprotein to the rate and efficiency seen in the original set of experiments. This interdependency may play a major role in determining the rate and completeness of processing of the precursors. In order to understand this interdependency a thorough knowledge of the type of factors within the gag and pol regions of the polyprotein precursors that influence processing must first be defined. In order to accomplish this goal the gag regions for the variants used in this research were sequenced. Based on the results of this sequence information, a comparison was made between gag regions that processed "normally" and

the gag region of the MD.03 variant that showed aberrant processing. Comparison of the sequences showed differences at three regions within the gag domain of the MD.03 variant when compared to that of the D2.21 variant: three amino acids within the C site, and two amino acids, one at 7 and one at 48 amino acids C-terminal of the C cleavage junction. Site directed mutagenesis was used to mutate the three amino acids within the C site to the amino acids found within the D2.21 variant. The gag region from this variant was shown to process to completion in both the original constructs and the chimeric precursors. It was thought, therefore, that reversion of these three amino acids would allow the gag region of the MD.03 variant to process to completion. This was not the case, however, as processing of this variant now containing the new C site was not significantly different from the processing phenotype observed for this variant before the changes were made.

The second change made to the MD.03 variant was a replacement of arginine residue with a lysine residue seen in the D2.21 variant. This change, in conjunction with the change in the C site region described above, was tested for the ability to affect a change in the observed processing phenotype. Again, there was no significant difference between the processing phenotype of this variant and that observed for the D2.21 variant. The final mutation made in this set of experiments was a conversion of the glutamic acid in the MD.03 variant to the aspartic acid seen in the D2.21 variant. The gag-pol precursor carrying this mutation, along with the two mutations described above, was expressed in the *E. coli*-based expression system. The results from this experiment, unlike that observed for the previous two experiments, showed the ability of the precursor to be

processed nearly to completion. To support these experiments, site directed mutagenesis was used to mutate both of the additional variant amino acids seen in this region without also first mutating the C site. Both of these mutants failed to convert the processing seen in this variant to that observed for the other variants. This is an interesting result in that all three mutations were necessary to restore processing of this gag region to that observed for the other variants. The need for all three mutations in order to restore "wild-type" processing points to a possible role for the three-dimensional structure of the substrate itself in influencing precursor processing. It is possible that residues within the gag region interact to form secondary structure and that this secondary structure is important in the control of precursor processing by the protease either by causing the cleavage sites to be more or less accessible to the PR. In variant MD.03, in which changes occurred within the gag region, this secondary structure could have formed incorrectly or not at all. This incorrect formation could occur as a result of the mutated residues within the C site causing a subtle change in PR binding of the site and therefore allowing the two remaining mutated residues to interact. If an incorrectly formed structure, such as a loop, formed an interaction with the active site residues of the enzyme or residues near the active site, this interaction could interfere with the ability of the PR to perform subsequent cleavage events. This interference could be the result of the loop structure remaining attached to the enzyme following cleavage at, in the case of variant MD.03, the C site, and thus blocking the active site of the enzyme. This model would explain the processing results obtained for Variant MD.03 in which cleavage of the C site was able to proceed, but subsequent processing of the precursor did not occur. These results add to

the possible number of factors that can affect processing of the gag pol polyprotein precursor.

In summary, this research has shown that numerous elements within the precursor can affect its own processing. These factors include; elements within the pol region, elements within the gag region, and a combination of elements from both the gag and pol regions that act together to influence processing. In addition, as was seen in chapter 6 of this dissertation, changes at multiple locations within the gag region can also affect polyprotein processing. This information is useful in defining the scope of possible changes within the precursors that can affect their processing.

Although the proteins used in this experiment were naturally arising, and as such were not influenced by protease inhibitors, the results of this research should still prove useful in understanding the range of variations within the pol and gag regions of the precursor used by the virus in order to escape drug pressure. Research by Zhang et. al (113) in which the gag region of the gag-pol precursor was sequenced following treatment with the protease inhibitor Indinavir, showed the presence of an amino acid change between the p7 and p1 proteins within the precursor, presumably as a result of the inhibitor used in the assay. In addition, two mutations, one at position 46 and one at position 82 were seen in the protease region of the precursor. When recombinant HIV-1, containing a protease with mutations at these two positions, were grown in cell culture there was a 68% reduction in its rate of replication when compared to that of wild-type virus. Introduction of the mutations seen at the gag p7/p1 cleavage site compensated for the defective protease gene. Similarly, rates of replication of viruses harboring protease

genes that contained mutations as a result of drug pressure were enhanced both in the presence and in the absence of the drug when paired with mutations at the p7/p1 and the p1/p6 gag cleavage sites. These results indicate the ability of the HIV-1 virus to change not only elements within the protease gene, but also elements within the gag region in order to ensure processing of the polyprotein precursors in the presence of inhibitors. In addition, these results lend support to those seen in this dissertation in which a combination of factors from both the pol and gag regions worked together to influence precursor processing. In addition to these findings, data obtained from the lab of Dr. Maureen Goodenow showed increasing variation within the C cleavage site of the gag-pol precursors of patients treated with protease inhibitors. These mutations seem to occur in conjunction with mutations within the protease gene itself that have been shown to confer resistance to protease inhibitors (unpublished results). These results, along with the results described here, point to the ability of the HIV-1 virus to modulate both elements within the gag and elements within the pol region of its precursors to compensate for a deficiency in another region of the precursor. In this manner the virus ensures the proper processing of the precursor even in the presence of potent antiretroviral therapy.

Future Directions

Much additional study is needed in order to understand fully the range and mutations used by the virus in order to ensure correct processing of the gag-pol polyprotein precursor. This research has shown that many factors, including a combination of factors originating from different regions of the precursor, are used by the virus to insure proper polyprotein precursor processing. In order to ensure that the

observations concerning precursor processing, as seen in the *E. coli* based expression system used in this study are similar to the precursor processing seen in a budding virion, expression of these variants in a mammalian cell line, in conjunction with an analysis of the observed cleavage products, is needed. Although the type of processing observed for the experiments described in this dissertation is similar to that observed using other expression systems, expression of variant gag-pol precursors as part of the intact virus within a mammalian cell line would add further support to the observations shown in this work providing similar results were obtained.

The second area of research that would further clarify the results of the work presented in this dissertation is the use of gag-pol regions from patients treated with HIV-1 protease inhibitors. As described previously, mutations within the gag region have been shown to compensate for a protease region that has diminished activity due to mutations caused by drug pressure. It would therefore be a straightforward matter to first test the ability of these precursors to process in their original form and then pair both of the individual domains with the opposite domain from a wild-type virus in order to observe the effects on the processing phenotype. This expression could also be performed with and without the use of potent protease inhibitors to determine if the mutations seen in the gag region are indeed necessary for the proper functioning of the protease under drug pressure. The results obtained from analysis of this expression should therefore determine whether the mutations seen in the gag region are indeed concomitant with those observed in the protease region. Understanding how the HIV-1 virus uses elements in the gag region to compensate for a protease whose activity has been diminished due to

drug pressure should prove useful in the development of future protease inhibitors.

In addition to expressing these variant gag-pol regions in a mammalian cell line, and using regions from gag-pol precursors isolated from patients being treated with protease inhibitors, knowledge of the structure of the gag-pol region prior to and during its processing by the protease would prove very beneficial. Structural information about the individual proteins making up the gag and pol region is known; however, structural information of intact gag-pol precursor or in intact gag precursor is currently not available. Being able to visualize the conformation of the precursor during the various stages of its processing could lead to the development of more effective inhibitors by determining which of the steps of the processing events are rate limiting and/or pivotal to the remaining processing events.

Structural information would also be of much use in determining how mutations in one domain compensate for and cooperate with changes in other domains. This information would be of particular interest for those variants that exhibit changes in both the pol and gag regions during antiretroviral therapy. Structural information would also prove useful in answering some of the questions raised during this work. As discussed earlier in this dissertation, mutations which appear in the gag region of variant MD.03 and indeed at the C site of this variant seem to affect cleavage at the A site. The A site is located some distance upstream of the residues which affect its processing. The ability of these residues to affect processing at this distance is not easily explained without a more thorough knowledge of the tertiary structure of the precursor during processing. Additionally, new classes of inhibitors, such as ones that inhibit the dimerization of the

gag-pol precursor and therefore prevent the formation of active protease could be developed using information obtained from a structural analysis of the precursor.

LIST OF REFERENCES

1. **Almog, N., R. Roller, G. Arad, E. Passi, M.A. Wainberg, and M. Kotler**. 1996. A p6Pol-protease fusion protein is present in mature particles of human immunodeficiency virus type 1. J Virol **70**:7228-7232.

2. **Arrigo, S.J. and K. Huffman**. 1995. Potent inhibition of human immunodeficiency virus type 1 (HIV-1) replication by inducible expression of HIV-1 PR multimers. J Virol **69**:5988-5994.

3. **Barre-Sinoussi, F., J.C. Chermann, F. Rey, M.T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier**. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220**:868-871.

4. **Barrie, K.A., E.E. Perez, S.L. Lamers, W.G. Farmerie, B.M. Dunn, J.W. Sleasman, and M.M. Goodenow**. 1996. Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. Virology **219**:407-416.

5. **Berger, E.A.** 1997. HIV entry and tropism: the chemokine receptor connection. AIDS **11**:3-16.

6. **Berkhout, B., R.H. Silverman, and K.T. Jeang**. 1989. Tat trans-activates the human immunodeficiency virus through a nascent RNA target. Cell **59**:273-282.

7. **Blanche, S., M. Tardieu, A. Duliege, C. Rouzioux, F. Le Deist, K. Fukunaga, M. Caniglia, C. Jacomet, A. Messiah, and C. Griscelli**. 1990. Longitudinal study of 94 symptomatic infants with perinatally acquired human immunodeficiency virus infection. Evidence for a bimodal expression of clinical and biological symptoms Am J Dis Child **144**:1210-1215.

8. **Blasig, C., C. Zietz, B. Haar, F. Neipel, S. Esser, N.H. Brookmeyer, E. Tschachler, S. Colombini, B. Ensoli, and M. Sturzl**. 1997. Monocytes in kaposi's sarcoma lesions are productively infected by human herpesvirus 8. J Virol **71**:7963-7968.

9. **Bonhoeffer, S., E.C. Holmes, and M.A. Nowak**. 1995. Causes of HIV diversity. Nature **376**:125

10. **Bugelski, P.J., R. Kirsh, and T.K. Hart**. 1994. HIV protease inhibitors: effects on viral maturation and physiologic function in macrophages. J Leukoc Biol **56**:374-380.

11. **Bukrinsky, M.I., N. Sharova, T.L. McDonald, T. Pushkarskaya, W.G. Tarpley, and M. Stevenson**. 1993. Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. Proc Natl Acad Sci U S A **90**:6125-6129.

12. **Camaur, D. and D. Trono**. 1996. Characterization of human immunodeficiency virus type 1 Vif particle incorporation. J Virol **70**:6106-6111.

13. **Cavert, W., D.W. Notermans, and K. Staskus**. 1996. Kinetics of response in lymphoid tissues to antiretroviral therapy of HIV-1 infection. Science **173**:960-964.

14. **Centers for Disease Control and Prevention**. 1996. HIV/AIDS Surveillance Report. Centers for Disease Control and Prevention **1**:3-33.

15. **Chen, Z., Y. Li, H.B. Schock, D. Hall, E. Chen, and L.C. Kuo**. 1995. Three-dimensional structure of a mutant HIV-1 protease displaying cross-resistance to all protease inhibitors in clinical trials. J Biol Chem **270**:21433-21436.

16. **Chun, T.W., L. Carruth, D. Finzi, X. Shen, J.A. DiGiuseppe, H. Taylor, M. Hermankova, K. Chadwick, J. Margolick, T.C. Quinn, Y.H. Kuo, R. Brookmeyer, M.A. Zeiger, P. Barditch-Crovo, and R.F. Siliciano**. 1997. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature **387**:183-188.

17. **Chun, T.W., D. Finzi, J. Margolick, K. Chadwick, M.D. Schwartz, and R.F. Siliciano**. 1995. In vivo fate of HIV-1-infected T cells: quantitative analysis of the transistion to stable latency. Nat Med **1**:1284-1290.

18. **Coffin, J.M., A.T. Haase, J.A. Levy, L. Montagnier, S. Oroszlan, N. Teich, K. Temin, T. Toyoshima, H. Varmas, P. Vogt, and R.A. Weiss**. 1998. Human immunodeficiency viruses. Science **232**:679

19. **Cohen, O.J., G. Pantaleo, and M. Holodniy**. 1996 Antiretroviral monotherapy in early stage human immunodeficiency virus desease has no detectable effect on virus load in peripheral blood and lymph nodes. J Infect.Dis. **173**:849-856.

20. **Condra, J.H., D.J. Holder, W.A. Schleif, O.M. Blahy, R.M. Danovich, L.J. Gabryelski, D.J. Graham, D. Laird, J.C. Quintero, A. Rhodes, H.L. Robbins, E. Roth, M. Shivaprakash, T. Yang, J.A. Chodakewitz, P.J. Deutsch, R.Y. Leavitt, F.E. Massari, J.W. Mellors, K.E. Squires, R.T. Steigbigel, H. Teppler, and E.A. Emini**. 1996. Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor. J Virol **70**:8270-8276.

21. **Condra, J.H., W.A. Schleif, O.M. Blahy, L.J. Gabryelski, D.J. Graham, J.C. Quintero, A. Rhodes, H.L. Robbins, E. Roth, and M. Shivaprakash**. 1995. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. Nature **374**:569-571.

22. **Curran, J., M. Essex, A. Fauci, V. DeVita, S. Hellan, and S. Rosenberg**. 1997. The Molecular Biology of HIV-1, AnonymousAIDS: Etiology, Diagnosis, Treatment, and Prevention. Lippincott-Raven, Philadelphia.

23. **Darke, P.L., S.P. Jordan, D.L. Hall, J.A. Zugay, J.A. Shafer, and L.C. Kuo**. 1994. Dissociation and association of the HIV-1 protease dimer subunits: equilibria and rates. Biochemistry **33**:98-105.

24. **Doms, R.W. and S.C. Peiper**. 1997. Unwelcomed guest with master keys: how HIV uses chemokine receptors for cellular entry. Virology **235**:179-190.

25. **Earl, P.L., R.W. Doms, and B. Moss**. 1990. Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. Proc Natl Acad Sci U S A **87**:648-652.

26. **Ellis, R.J., S.A. Spector, and K. Hsia**. 1998. HIV-1 viral dynamics differ in cerebrospinal fluid and plasma. 5th conference on Retroviruses and Opportunistic Infections (Abstract 243)

27. **Embretson, J., M. Zupancic, and J.L. Ribas**. 1993. Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS. Nature **362**:359-362.

28. **Epstein, L.G. and H.E. Gendelman**. 1993. Human immunodeficiency virus type 1 infection of the nervous system: pathogenetic mechanisms. Ann Neurol. **33**:429-436.

29. **Erickson, J., D.J. Neidhart, J. VanDrie, D.J. Kempf, X.C. Wang, D.W. Norbeck, J.J. Plattner, J.W. Rittenhouse, M. Turon, and N. Wideburg**. 1990. Design, activity, and 2.8 A crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease. Science **249**:527-533.

30. **Erickson, V., J. Manfredi, P. Viitanen, D.E. Tribe, R. Tritch, C.A. Hutchison, D.D. Loeb, and R. Swanstrom**. 1989. Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease. AIDS Res Hum Retroviruses **5**:577-591.

31. **Feinberg, M.B., R.F. Jarret, A. Aldovini, R.C. Gallo, and F. Wong-Staal**. 1986. HTLV-III expression and production envolves complex regulation at the levels of splicing and translation of viral RNA. Cell **46**:807-817.

32. **Finzi, D., M. Hermankova, T. Pierson, L.M. Carruth, C. Buck, R.E. Chaisson, T.C. Quinn, K. Chadwick, J. Margolick, R. Brookmeyer, J. Gallant, M. Markowitz, D.D. Ho, D.D. Richman, and R.F. Siliciano**. 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science **278**:1295-1300.

33. **Freed, E.O., J.M. Orenstein, W. Buckler, and M.A. Martin**. 1994. Single amino acid changes in the human immunodeficiency virus type 1 matrix protein block virus particle production. J Virol **68**:5311-5320.

34. **Gabuzda, D. and M.S. Hirsh**. 1987. Neurologic manifestations of infection with human immunodeficiency virus: clincal features and pathogenesis. Ann Intern Med. **107**:383-391.

35. **Gallaher, W.R.** 1987. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. Cell **50**:327-328.

36. **Gallo, R.C., S.Z. Salahuddin, M. Popovic, G.M. Shearer, M. Kaplan, B.F. Haynes, T.J. Palker, R. Redfield, J. Oleske, and B. Safai**. 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science **224**:500-503.

37. **Gallo, R.C., P.S. Sarin, E.P. Gelmann, M. Robert-Guroff, E. Richardson, V.S. Kalyanaraman, D. Mann, G.D. Sidhu, R.E. Stahl, S. Zolla-Pazner, J. Leibowitch, and M. Popovic**. 1983. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science **220**:865-867.

38. **Gelderblom, H.R. and E.H.H.S.W.T.P.G. Ozel**. 1998. Fine structure of human immunodeficiency virus (HIV), immunolocalization of structural proteins and virus-cell relation. Micron Microscop **19**:41-60.

39. **Gonzalez-Scarano, F., M.N. Waxham, A.M. Ross, and J.A. Hoxie**. 1987. Sequence similarities between human immunodeficiency virus gp41 and paramyxovirus fusion proteins. AIDS Res Hum Retroviruses **3**:245-252.

40. **Greene, W.C.** 1991. The molecular biology of human immunodeficiency virus type 1 infection. N Engl J Med **324**:308-317.

41. **Haase, A.T.** 1986. Pathogenesis of lentivirus infections. Nature **322**:130-136.

42. **Ho, D.D.** 1995. HIV-1 dynamics in vivo. J Biol Regul.Homeost.Agents **9**:76-77.

43. **Ho, D.D., A.U. Neumann, A.S. Perelson, W. Chen, J.M. Leonard, and M. Markowitz**. 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature **373**:123-126.

44. **Hong, S.S. and P. Boulanger**. 1993. Assembly-defective point mutants of the human immunodeficiency virus type 1 Gag precursor phenotypically expressed in recombinant baculovirus-infected cells. J Virol **67**:2787-2798.

45. **Kaczorowski, S., M. Ochocka, M. Kaczorowska, R. Aleksandrowicz, M. Matysiakl, and M. Karwacki**. 1995. Is P-glycoprotein a sufficient marker for multidrug resistance in vivo? immunohistochemical staining for P-glycoprotein in children and adult leukemia: correlation with clinical outcome. Leuk Lymphoma **20**:143-152.

46. **Kaplan, A.H., M. Manchester, and R. Swanstrom**. 1994. The activity of the protease of human immunodeficiency virus type 1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. J Virol **68**:6782-6786.

47. **Kaplan, A.H., J.A. Zack, M. Knigge, D.A. Paul, D.J. Kempf, D.W. Norbeck, and R. Swanstrom**. 1993. Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. J Virol **67**:4050-4055.

48. **Kim, R.B., M.F. Fromm, and C. Wandel**. 1998. The drug transporter P-glycoproteinlimits oral absorption and brain entry of HIV-1 protease inhibitors. J Clin Invest **101**:289-294.

49. **Kohl, N.E., E.A. Emini, W.A. Schleif, L.J. Davis, J.C. Heimbach, R.A. Dixon, E.M. Scolnick, and I.S. Sigal**. 1988. Active human immunodeficiency virus protease is required for viral infectivity. Proc Natl Acad Sci U S A **85**:4686-4690.

50. **Krausslich, H.G., R.H. Ingraham, M.T. Skoog, E. Wimmer, P.V. Pallai, and C.A. Carter**. 1989. Activity of purified biosynthetic proteinase of human immunodeficiency virus on natural substrates and synthetic peptides. Proc Natl Acad Sci U S A **86**:807-811.

51. **Kure, K., Llena J.F., and W.D. Lyman**. 1991. Human immunodeficiency virus-1 infection of the nervous system: an autopsy study of 268 adult, pediatric, and fetal brains. Hum Pathol. **22**:700-710.

52. **Lam, P.Y., P.K. Jadhav, C.J. Eyermann, C.N. Hodge, Y. Ru, L.T. Bacheler, J.L. Meek, M.J. Otto, M.M. Rayner, and Y.N. Wong**. 1994. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. Science **263**:380-384.

53. **Le Grice, S.F., R. Ette, J. Mills, and J. Mous**. 1989. Comparison of the human immunodeficiency virus type 1 and 2 proteases by hybrid gene construction and trans-complementation. J Biol Chem **264**:14902-14908.

54. **Le Grice, S.F., J. Mills, and J. Mous**. 1988. Active site mutagenesis of the AIDS virus protease and its alleviation by trans complementation. Embo J **7**:2547-2553.

55. **Lee, P.P. and M.L. Linial**. 1995. Inhibition of wild-type HIV-1 virus production by a matrix deficient Gag mutant. Virology **208**:808-811.

56. **Leigh Brown, A.J. and E.C. Holmes**. 1994. Evolutionary biology of human immunodeficiency virus. Annu Rev Ecol Syst **25**:127-165.

57. **Leis, J., D. Baltimore, J.M. Bishop, J.M. Coffin, E. Fleissner, S.P. Goff, S. Oroszlan, H. Robinson, A.M. Skalka, H.M. Temin, and V. Vogt**. 1988. Standardized and simplified nomenclature for proteins common to all retroviruses. J Virol **62**:1808-1809.

58. **Levy, J.A.** 1986. The multifaceted retrovirus. Cancer Res **46**:5457-5468.

59. **Levy, J.A., A.D. Hoffman, S.M. Kramer, J.A. Landis, J.M. Shimabukuro, and L.S. Oshiro**. 1984. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science **225**:840-842.

60. **Liu, H., X. Wu, M. Newman, G.M. Shaw, B.H. Hahn, and J.C. Kappes**. 1995. The Vif protein of human and simian immunodeficiency viruses is packaged into virions and associates with viral core structures. J Virol **69**:7630-7638.

61. **Louis, J.M., N.T. Nashed, K.D. Parris, A.R. Kimmel, and D.M. Jerina**. 1994. Kinetics and mechanism of autoprocessing of human immunodeficiency virus type 1 protease from an analog of the Gag-Pol polyprotein. Proc Natl Acad Sci USA **91**:7970-7974.

62. **Lu, Y.L., P. Spearman, and L. Ratner**. 1993. Human immunodeficiency virus type 1 viral protein R localization in infected cells and virions. J Virol **67**:6542-6550.

63. **Luban, J., C. Lee, and S.P. Goff**. 1993. Effect of linker insertion mutations in the human immunodeficiency virus type 1 gag gene on activation of viral protease expressed in bacteria. J Virol **67**:3630-3634.

64. **Luukkonen, B.G., E.M. Fenyo, and S. Schwartz**. 1995. Overexpression of human immunodeficiency virus type 1 protease increases intracellular cleavage of Gag and reduces virus infectivity. Virology **206**:854-865.

65. **Malim, M.H., J. Hauber, S.Y. Le, J.V. Maizel, and B.R. Cullen**. 1989. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature **338**:254-257.

66. **Malim, M.H., L.S. Tiley, D.F. McCarn, J.R. Rusche, J. Hauber, and B.R. Cullen**. 1990. HIV-1 structural gene expression requires binding of the Rev trans- activator to its RNA target sequence. Cell **60**:675-683.

67. **Mann, J. and D. Tarantola**. 1996. Aids in the World II. Oxford University Press, New York.

68. **Markowitz, M., H. Mo, D.J. Kempf, D.W. Norbeck, T.N. Bhat, J.W. Erickson, and D.D. Ho**. 1995. Selection and analysis of human immunodeficiency virus type 1 variants with increased resistance to ABT-538, a novel protease inhibitor. J Virol **69**:701-706.

69. **Marlink, R.** 1996. Lessons from the second AIDS virus, HIV-2. AIDS **10**:689-699.

70. **McCune, J.M., L.B. Rabin, M.B. Feinberg, M. Lieberman, J.C. Kosek, G.R. Reyes, and I.L. Weissman**. 1988. Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. Cell **53**:55-67.

71. **Montagnier, L., A.G. Doyle, F. Barre-Sinoussi, S. Chamaret, J. Gruest, M.T. Nugeyre, F. Rey, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, A.G. Saimot, W. Rozenbaum, J.C. Gluckman, D. Klatzman, E. Vilmer, C. Griscelli, C. Gazengel, and J.B. Brunet**. 1984. A new human T-lymphotropic retrovirus: characterization and possible role in lymphadenopathy and acquired immune deficiency syndromes, p. 363-379. In L. Gross (ed.), Human T-Cell Leukemia/Lymphoma Virus. Cold Spring Harbor Laboratory, Cold Spring Harbor.

72. **Myers, G., B.H. Hahn, J.W. Mellors, F.E. McCutcheon, G.N. Pavlakis, L.E. Henderson, B. Korber, and K.T. Jeang**. 1995. Human Retroviruses and AIDS 1995: A compilation and analysis of nucleic acid and amino acid sequences.

73. **Myers, G., K. MacInnes, and B. Korber**. 1992. The emergence of simian/human immunodeficiency viruses. AIDS Res Hum Retroviruses **8**:373-386.

74. **Nabel, G.J. and D. Baltimore**. 1987. An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature **326**:711-713.

75. **Orenstein, J.M., C. Fox, and S.M. Wahl**. 1997. Macrophages as a source of HIV during oppurtunistic infections. Science **276**:1857-1861.

76. **Ou, S.H.I. and R.B. Gaynor**. 1995. The Retroviridae. p.97-187. Plenum Press, New York.

77. **Partin, K., H.G. Krausslich, L. Ehrlich, E. Wimmer, and C. Carter**. 1990. Mutational analysis of a native substrate of the human immunodeficiency virus type 1 proteinase. J Virol **64**:3938-3947.

78. **Pauza, C.D.** 1990. Two bases are deleted from the termini of HIV-1 linear DNA during integrative recombination. Virology **179**:886-889.

79. **Perelson, A.S., P. Essunger, and Y. Cao**. 1997. Decay characteristics of HIV-1 infected compartments during combination therapy. Nature **387**:188-191.

80. **Perelson, A.S., A.U. Neumann, M. Markowitz, J.M. Leonard, and D.D. Ho**. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science **271**:1582-1586.

81. **Peterlin, B.M.** 1995. Molecular Biology of HIV. p.185-238. Plenum Press, New York.

82. **Pettit, S.C., M.D. Moody, R.S. Wehbie, A.H. Kaplan, P.V. Nantermet, C.A. Klein, and R. Swanstrom**. 1994. The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. J Virol **68**:8017-8027.

83. **Pettit, S.C., J. Simsic, D.D. Loeb, L. Everitt, C.A. Hutchison, and R. Swanstrom**. 1991. Analysis of retroviral protease cleavage sites reveals two types of cleavage sites and the structural requirements of the P1 amino acid. J Biol Chem **266**:14539-14547.

84. **Picker, L.J. and M.H. Siegelman**. 1993. Lymphoid tissues and organs, p. 145-197. In W.E. Paul (ed.), Fundamental Immunology. Raven Press Ltd., New York.

85. **Pomerantz, R.J., D.R. Kuritzkes, and S.M. de la Monte**. 1987. Infection of the retina by human immunodeficiency virus type I. N Engl J Med **317**:1643-1647.

86. **Purcell, D.F. and M.A. Martin**. 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol **67**:6365-6378.

87. **Quinn, T.C.** 1996. Global burden of the HIV pandemic published erratum appears in Lancet 1996 Jul 27;348(9022):276]. Lancet **348**:99-106.

88. **Roberts, N.A., J.A. Martin, D. Kinchington, A.V. Broadhurst, J.C. Craig, I.B. Duncan, S.A. Galpin, B.K. Handa, J. KAY, and A. Krohn**. 1990. Rational design of peptide-based HIV proteinase inhibitors. Science **248**:358-361.

89. **Rose, J.R., L.M. Babe, and C.S. Craik**. 1995. Defining the level of human immunodeficiency virus type 1 (HIV-1) protease activity required for HIV-1 particle maturation and infectivity. J Virol **69**:2751-2758.

90. **Sardana, V.V., A.J. Schlabach, P. Graham, B.L. Bush, J.H. Condra, J.C. Culberson, L. Gotlib, D.J. Graham, N.E. Kohl, and R.L. LaFemina**. 1994. Human immunodeficiency virus type 1 protease inhibitors: evaluation of resistance engendered by amino acid substitutions in the enzyme's substrate binding site. Biochemistry **33**:2004-2010.

91. **Sawai, E.T., A. Baur, H. Struble, B.M. Peterlin, J.A. Levy, and C. Cheng-Mayer**. 1994. Human immunodeficiency virus type 1 Nef associates with a cellular serine kinase in T lymphocytes. Proc Natl Acad Sci U S A **91**:1539-1543.

92. **Schecter, I. and B. Berger**. 1967. On the size of the active site in proteases. Biochem Biophys Res Commun **27**:157-162.

93. **Schlegel, P.N. and S.K. Chang**. 1998. Physiology of male reproduction: the testes, epididymis, and ductus deferens, p. 1254-1286. In P.C. Walsh, A.B. Retik, E.D. Vaughan, and A.J. Wein (eds.), Campbell's Urology. WB Saunders Co., Philadelphia.

94. **Schock, H.B., V.M. Garsky, and L.C. Kuo**. 1996. Mutational anatomy of an HIV-1 protease variant conferring cross-resistance to protease inhibitors in clinical trials. Compensatory modulations of binding and activity. J Biol Chem **271**:31957-31963.

95. **Schwartz, S., B.K. Felber, and G.N. Pavlakis**. 1992. Mechanism of translation of monocistronic and multicistronic human immunodeficiency virus type 1 mRNAs. Mol Cell Biol **12**:207-219.

96. **Seelmeier, S.S.H., V. Turk, and K. von der Helm**. 1988. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. Proc Natl Acad Sci U S A **85**:6612-6616.

97. **Sodroski, J., W.C. Goh, C. Rosen, A. Dayton, E. Terwilliger, and W. Haseltine**. 1986. A second posttranscriptional transactivator gene required for HTLV-III expression. Nature **321**:412-417.

98. **Srinivasakumar, N., M.L. Hammarskjold, and D. Rekosh**. 1995. Characterization of deletion mutations in the capsid region of human immunodeficiency virus type 1 that affect particle formation and Gag-Pol precursor incorporation. J Virol **69**:6106-6114.

99. **Stahle, L., C. Martin, J.O. Svensson, and A. Sonnerborg**. 1997. Indinavir in cerebrospinal fluid of HIV-1 infected patients. Lancet **350**:1823-1823.

100. **Steimer, K.S., J.P. Puma, M.D. Power, M.A. Powers, N. George, J.C. Stephans, J.A. Levy, P. Sanchez, P.A. Luciw, and P.J. Barr**. 1986. Differential antibody responses of individuals infected with AIDS-associated retroviruses surveyed using the viral core antigen p25gag expressed in bacteria. Virology **150**:283-290.

101. **Stevenson, M.** 1994. Identification of factors that govern HIV replication in non-dividing host cells. AIDS Res Hum Retroviruses **10**:254-258.

102. **Tang, J. and R.N.S. Wong**. 1987. A structural comparison of a group of Aspartic Proteinases. J Cell Biochem **33**:53-63.

103. **Tisdale, M., R.E. Myers, B. Maschera, N.R. Parry, N.M. Oliver, and E.D. Blair**. 1995. Cross-resistance analysis of human immunodeficiency virus type 1 variants individually selected for resistance to five different protease inhibitors. Antimicrob Agents Chemother **39**:1704-1710.

104. **Toh, H., M. Ono, K. Saigo, and T. Miyata**. 1985. Retroviral gag and DNA endonuclease coding sequences in IgE-binding factor gene. Nature **318**:388-389.

105. **Tossi, A., N. Antcheva, D. Romeo, and S. Miertus**. 1995. Development of pseudopeptide inhibitors of HIV-1 aspartic protease: analysis and tuning of the subsite specificity. Pept Res **8**:328-334.

106. **Tritch, R.J., Y.E. Cheng, F.H. Yin, and V. Erickson**. 1991. Mutagenesis of protease cleavage sites in the human immunodeficiency virus type 1 gag polyprotein. J Virol **65**:922-930.

107. **Tyagi, S.C., S.R. Simon, and C.A. Carter**. 1994. Effect of pH and nonphysiological salt concentrations on human immunodeficiency virus-1 protease dimerization. Biochem Cell Biol **72**:175-181.

108. **Wang, C.T. and E. Barklis**. 1993. Assembly, processing, and infectivity of human immunodeficiency virus type 1 gag mutants. J Virol **67**:4264-4273.

109. **Welker, R., H. Kottler, H.R. Kalbitzer, and H.G. Krausslich**. 1996. Human immunodeficiency virus type 1 Nef protein is incorporated into virus particles and specifically cleaved by the viral proteinase. Virology **219**:228-236.

110. **Wlodawer, A. and J. Erickson**. 1993. Structure based inhibitors of HIV-1 protease. Annu Rev Biochem **62**:543-585.

111. **Yu, X., Z. Matsuda, Q.C. Yu, T.H. Lee, and M. Essex**. 1993. Vpx of simian immunodeficiency virus is localized primarily outside the virus core in mature virions. J Virol **67**:4386-4390.

112. **Yuan, X., X. Yu, T.H. Lee, and M. Essex**. 1993. Mutations in the N-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor. J Virol **67**:6387-6394.

113. **Zhang, Y.M., H. Imamichi, T. Imamichi, H.C. Lane, J. Falloon, M.B. Vasudevachari, and N.P. Salzman**. 1997. Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. J Virol **71** :6662-6670.

114. **Zybarth, G. and C. Carter**. 1995. Domains upstream of the protease (PR) in human immunodeficiency virus type 1 Gag-Pol influence PR autoprocessing. J Virol **69**:3878-3884.

115. **Zybarth, G., H.G. Krausslich, K. Partin, and C. Carter**. 1994. Proteolytic activity of novel human immunodeficiency virus type 1 proteinase proteins from a precursor with a blocking mutation at the N terminus of the PR domain. J Virol **68**:240-250.

BIOGRAPHICAL SKETCH

Gregory Charles Bloom was born on December 29th 1966 in Jeffersonville, Indiana. He was raised, along with his two siblings Connie and Chris, by his parents Bernard Charles Bloom and Lorane Sue Bloom in the town of Scottsburg. He attended primary and middle school in Scott County. After graduating from Scottsburg High School in 1985, he spent two years in the United States Army, 18 months of which were in Kaiserslautern West Germany. Here he met his wife of 11 years, Annemarie.

Gregory graduated from Indiana University in 1993 with a Bachelor of Arts degree with his major field of study being biology. During his undergraduate years, Gregory worked as a lab assistant under the guidance of Dr. Gretchen Kirchner, and Dr. David W. Taylor. It was during this time he developed his love of biochemistry and molecular biology.

Gregory has been fortunate enough to pursue his goal of obtaining a Ph.D. under the guidance of Dr. Ben M. Dunn. Dr. Dunn has been of invaluable assistance in helping Gregory obtain his degree, and his hard work and guidance toward this goal is greatly appreciated. Following completion of his doctoral studies Gregory plans to pursue a career in research in the area of drug discovery.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ben M. Dunn, Chair
Distinguished Professor of Biochemistry and Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Charles Allen
Professor of Biochemistry and Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Nancy Denslow
Associate Scientist of Biochemistry and Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Harry Nick
Professor of Neuroscience

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Maureen Goodenow
Associate Professor of Pathology, Immunology, and Laboratory Medicine

This dissertation was submitted to the Graduate Faculty of the College of Medicine and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1998

Dean, College of Medicine

Dean, Graduate School